THE UNIVERSITY
of EDINBURGH

East Coast Fever and vaccination at the livestock / wildlife interface

Fiona Katharine Allan

Submitted for the degree of Doctor of Philosophy

University of Edinburgh

2019

Haba na haba hujaza kibaba - Little by little fills up the measure

(Swahili proverb)

Declaration

I declare that the work presented in this thesis is my own original work, except where specified. No part of this work has been, or will be submitted for any other degree, or professional qualification.

Fiona Katharine Allan

June 2019

Abstract

East Coast fever (ECF) in cattle is caused by the tick-borne protozoan *Theileria parva*, and is transmitted by the three-host tick *Rhipicephalus appendiculatus.* The African buffalo (*Syncerus caffer*) is the natural host for *T. parva* but does not suffer disease, whereas ECF is often fatal in cattle, causing annual losses of more than $300 million. A live vaccine is available, the Infection and Treatment Method (ITM) Muguga Cocktail, but antigenic diversity of the parasite - particularly of buffalo-derived *T. parva* - results in variable protection.

The project aimed to determine the prevalence and genetic and antigenic diversity of the *T. parva* population in cattle and buffalo, in an area adjacent to the Serengeti National Park (SENAPA), Tanzania, where livestock co-exist with buffalo, as well as ascertaining herd vector control practices in the study area, to help inform future control strategies.

Field data were generated by designing and implementing a randomised cross-sectional sampling survey. Cattle were sampled (n=770) from 48 herds and blood samples analysed by diagnostic nested p104 PCR to establish a *T. parva* prevalence of 5.07% (CI: 3.70-7.00%). This prevalence was lower than in other hyperendemic areas. Half body tick counts were recorded on every cow and although 100% farmers reported seeing ticks on their cattle, tick counts were very low with 78% cattle having zero ticks. A questionnaire survey was created and carried out with 120 farmers, including the 48 sampled herds, to obtain data on vector control. Questionnaire data indicated significant use of acaracide with 79% (CI: 71-85%) of farmers spraying and 41% (32-49%) dipping cattle routinely. Some farmers reported very frequent spraying, as often as every four days. All acaricides used were from the same synthetic pyrethroid drug class, cypermethrin. Local workshops were held to discuss findings and validate results. These data indicate high levels of acaricide use, which may be responsible for the low observed tick burdens and low *T. parva* prevalence. The vector control is farmer-led and aimed at both ticks and tsetse flies. The levels of acaricide use raise concerns regarding sustainability, as large scale use of a single acaricide compound clearly represents a risk for the

development of resistance.

A genotyping pipeline was designed to characterise genetic diversity in *T. parva* field samples. Previous studies have shown that a number of *T. parva* antigens are recognised by $CD8^+$ T cells in immunised cattle, with several of these antigens demonstrating polymorphism, with greater allelic diversity in buffalo-derived *T. parva* than in cattle-derived parasite populations. A panel of twelve antigen-encoding genes was investigated, with successful species-specific full length or near-full length amplification of four genes – Tp1, A14, Tp4 and N60. A panel of DNA of known parasite composition stocks, including diverse *Theileria* species and multiple characterised isolates of both *T. parva* and the closely related *T.* sp. (buffalo), was used to validate specificity and sensitivity of primers before applying them to DNA from cattle and buffalo samples from the Ol Pejeta game conservancy, Kenya. The Ol Pejeta site had a known grazing history; buffalo are endemic and no cattle had grazed there for several years and so the infections detected in cattle that were introduced to the area would all be buffalo-derived. This setting allowed for comparison of diversity of *T. parva* circulating in buffalo with that of parasites acquired by sentinel cattle. Amplicons of antigen genes Tp1, A14 and N60, from two cattle and two buffalo, were sequenced by PacBio long-read technology (RSII platform). Cluster analysis showed that diversity was greater in buffalo parasite populations compared to cattle parasite populations from the Ol Pejeta conservancy in Tp1 and N60, but there were shared parasite populations in cattle and buffalo for all three genes. Diversity was greatest in Tp1 which had 109 clusters (13 cattle, 55 buffalo, 41 shared), A14 showed little diversity, with 23 clusters (10 cattle, 0 buffalo, 13 shared). N60 had 37 clusters (2 cattle, 4 buffalo, 37 shared).

After validation of the pipeline using samples from Ol Pejeta, DNA from the Serengeti cross-sectional cattle samples (n = 770), as well as cattle samples from several other timepoints and locations in the SENAPA study area (n = 832), and buffalo samples (n = 22) from the SENAPA study area was analysed for genetic and antigenic diversity of *T. parva*. Samples positive for *T. parva*, by p104 nPCR (149 cattle, 22 buffalo), were used to amplify antigen genes

Tp1 and N60. PacBio long-read sequencing (Sequel platform) was applied to multiplexed amplicons. These data showed overall that allelic diversity was significantly greater in buffalo-derived parasites compared to cattle-derived, with some alleles shared between buffalo-derived and cattle-derived parasites. Both Tp1 and N60 showed a high degree of polymorphism at the nucleotide level. For Tp1, there were 97 variable loci over the 1618 bp length, resulting in 86 gene allele variants. For N60, there were 14 variable loci over the 983 bp length resulting in 216 gene allele variants. Most alleles were unique to buffalo (55/86 in Tp1; 191/216 in N60), with a smaller proportion unique to cattle (19/86 in Tp1; 14/216 in N60) and relatively few shared (12/86 in Tp1; 11/216 in N60). At the amino acid level, Tp1 showed a large number of variants resulting in four epitope variants observed. In contrast, all nucleotide polymorphisms in N60 were synonymous and this gene was completely conserved at the amino acid level. Sequences identical to the genome reference strain and vaccine component, *T. parva* Muguga (TpM), were identified in several individuals; 9 cattle and 1 buffalo for Tp1, and 5 cattle and 1 buffalo for N60. There were indications of substructuring, by both genetic distance and phylogenetics, with most alleles found in cattle being closely related to the TpM reference allele, and the most distantly related cluster mainly being found only in buffalo. This, however, requires further sampling and increased numbers to confirm. These results have implications for ITM vaccine efficacy in the study area as well as demonstrating N60 as a potential candidate gene for alternative vaccines where cattle and buffalo interact.

Lay summary

East Coast Fever (ECF) is a disease of cattle that is caused by a blood-dwelling parasite called *Theileria parva*. *Theileria parva* parasites are transmitted to cattle by a tick called *Rhipicephalus appendiculatus*, also known as the 'brown ear tick'. The disease is present in eastern, central and southern parts of Africa and the natural host is the African buffalo (*Syncerus caffer*). The buffalo host does not show signs of disease, as it is thought to have adapted to the *Theileria parva* parasite over a long evolutionary period. In cattle, however, the disease is often fatal and causes annual losses of more than $300 million. There is a live vaccine available called the Infection and Treatment (ITM) Muguga Cocktail, but antigenic diversity of the parasite – particularly of parasites from buffalo – results in variable protection.

The project aimed to 1) determine the prevalence of *Theileria parva* in cattle and buffalo in an area adjacent to the Serengeti National Park (SENAPA), Tanzania, where cattle co-exist with buffalo, 2) determine the genetic and antigenic diversity of the *Theileria parva* population in the study area, and 3) establish the tick control practices that were being used by farmers in the study area.

Data were generated by collecting blood samples from 770 cattle, in 48 herds. Blood samples were analysed by molecular analyses to detect DNA and identify the *Theileria parva* parasite; this indicated a prevalence of 5.07% (CI: 3.70-7.00%) (39 of 770 cattle positive for *Theileria parva*). This prevalence is lower than in other areas of Tanzania where ECF is present. Tick counts were recorded on every cow and although farmers reported seeing ticks on their cows, the tick counts were very low; 78% cattle had zero ticks. A questionnaire was created and carried out with 120 farmers, including the 48 herds that had been sampled, in order to establish the tick control practices that farmers were using. The questionnaire data indicated that farmers were using large amounts of tick prevention products (known as acaricides); 79% (CI: 71-85%) of farmers were spraying their cattle with acaricide and 41% (CI: 32-49%) were dipping their cattle. Some farmers reported spraying their cattle very often,

sometimes every four days. All of the acaricide products that the farmers were using were from the same pyrethroid drug class called cypermethrin. Workshops were held to discuss findings with the farmers. These data indicate high levels of acaricide use, which may explain the low tick counts on the cattle and the low prevalence of *Theileria parva*. The practice of using acaricides was decided by the farmers, to prevent both ticks and tsetse flies (the vector of another disease affecting cattle, African animal trypanosomiasis). There is concern for how sustainable the practice of acaricide use is because ticks are likely to develop resistance to the products due to frequent exposure.

In order to determine the genetic and antigenic diversity of *Theileria parva* in the study area, a molecular 'pipeline' was designed. Antigens are molecules on the surface of infections that the immune system recognises as foreign. The immune system makes antibodies to fight off infection. Previous studies have shown that a number of *Theileria parva* antigens are recognised by particular cells, CD8 T cells, in immunised cattle. Several of these antigens have been shown to have a lot of genetic variation, called polymorphism. It has been shown that there is more genetic variation in *Theileria parva* in buffalo than in cattle parasite populations. A panel of twelve genes that encode antigens was investigated and four gene were amplified successfully. Using the PCR method to identify DNA, four genes were successfully amplified - Tp1, A14, Tp4 and N60. Short genetic sequences called primers were designed and trialled on DNA of known laboratory strains of *Theileria parva* and similar parasites to check their specificity for *Theileria parva*. As a trial, DNA samples from the Ol Pejeta game conservancy, Kenya, were used because this area had a known grazing history; no cattle had grazed there for several years and so the infections in cattle would be from buffalo (buffalo-derived). This setting allowed for comparison of diversity of *Theileria parv*a in buffalo with that of parasites from cattle. The genetic sequences of three genes, Tp1, A14 and N60 were sequenced. Sequences were clustered by similarity and showed that diversity was greater in buffalo parasite populations compared to cattle from Ol Pejeta, but there were shared parasites in both cattle and buffalo.

After this trial to validate the pipeline, DNA was used from the cattle samples collected in the Serengeti sampling survey (770), as well as other samples from other timepoints and locations in the study area (832) and buffalo samples (22) from the study area to investigate diversity in *Theileria parva*. Samples were screened by PCR to diagnose *Theileria parva* infection and those that were positive were used to amplify genes Tp1 and N60. These data showed overall that there was more diversity in buffalo-derived parasites compared to cattle-derived, with some sharing between both. Both Tp1 and N60 showed high levels of genetic diversity, with 86 variants of Tp1 and 216 variants of N60. Genetic sequences make up proteins and the sequences of Tp1 resulted in four variants of the Tp1 protein (epitope variants) whereas there was no variation in the N60 protein. In some of the cattle and buffalo, sequences were found that were identical to one of the components of the ITM vaccine, called *Theileria parva* Muguga (TpM). Evolutionary relationships were examined amongst the genetic variants in cattle and buffalo and the variants found in cattle tended to group with the TpM sequence while the buffalo grouped further away. These results have implications for use of the ITM vaccine in the study area and show that N60 could be a possible candidate for alternative vaccine studies because it has no antigenic diversity and could protect against all parasite strains.

Acknowledgements

I am indebted to the support and assistance of many wonderful people, without whom this project would not have been possible. Enormous thanks to my supervisors Professor Ivan Morrison, Dr Liam Morrison, Dr Harriet Auty and Dr Ian Handel, for their awe-inspiring knowledge, expertise and patience, and for taking a leap of faith in selecting a mature veterinarian straight out of clinical practice for this studentship.

Professor Ivan Morrison, your contribution to research in the field of *Theileria parva* is outstanding and I am most grateful to you for imparting a fraction of your knowledge to me. Dr Liam Morrison, thank you for your day-to-day support and supervision, trouble-shooting when lab work was particularly challenging, your ever clear vision of the project and encouragement to believe in my own ability; I am incredibly grateful to you. I have learnt so much from travelling and working in the field with Dr Harriet Auty; your passion for and knowledge of the Serengeti is an inspiration. I hope I will get back to Seronera and Mugumu again… Particular thanks to Dr Ian Handel; roped in as an 'extra' supervisor, your patient tutelage in R software statistics and general pastoral care and encouragement have been a godsend.

I am enormously grateful to BBSRC and GALVmed, in particular Dr Jeremy Salt, for funding this PhD.

Thank you to my thesis committee, Associate Professor Anagha Joshi (now University of Bergen, Norway) and Professor Lonneke Vervelde, for your support and continued interest in my project over the past four years.

It has been a pleasure and a privilege to be a member of the cheerful and dynamic Morrison lab group for the past four years and I consider all the group members, past and present, friends as well as colleagues; Omar Alfituri, Ruben Barroso, Stephen Chiweshe, Carly Hamilton, James Iremonger, Siddharth Jayaraman, Edith Paxton, Jessica Powell, Valentina Riggio, Pieter Steketee, Laura Vezza and Rachel Young – thank you all. Enormous thanks

go to Siddharth Jayaraman for your patience, time and persistence in sharing your bioinformatics knowledge and skills – thank you.

I have had wonderful assistance and support provided by several special and hard-working colleagues at the Liverpool School of Tropical Medicine (LSTM), namely Professor Steve Torr, Dr Jennifer Lord and Rachel Lea, my wonderful room-mate on many a field trip. Huge thanks also go to my Tanzanian colleagues, namely Emmanuel Sindoya, Ahmed Lugelo, Dr Geofrey Mbata, Dr Furaha Mramba, Dr Metchilda Byamungu and Kunda Mnzava. My time spent in Tanzania was exceptionally special and I feel so fortunate to have worked with you all. Asante sana.

Dr Alan Walker, retired entomologist and enthusiast of all natural things, provided me with expert training when preparing for field work and gave his time and enthusiasm so willingly, for which I am most grateful.

Samples and advice were kindly provided from colleagues at the International Livestock Research Institute (ILRI), Kenya, and I gratefully thank Dr Phil Toye, Dr Hanneke Hemmink, Dr Annie Cook and Maurine Chepkwony. Thanks also to Tara Sheldrake for providing cell lines.

Thank you to wonderful Liz Archibald, for administrative support and advice since day one, and also to Marshall Dozier, for fixing a variety of formatting ‘issues’ over the years.

Thank you to all the many friends I have made during my time spent at The Roslin Institute, in particular Marianne Keith, Marie O’Shea and Prerna Vohra, and to Edith Paxton I owe a debt of gratitude; you taught me all I know in the lab and you are not only a wonderful colleague but a dear friend too. I am also hugely thankful for the friendship and support from old friends, in particular Mary McCulloch with whom I shared some of this journey; your understanding of the PhD process, along with endless cups of tea, have been such a source of wisdom and comfort.

This thesis is in memory of my Great Uncles, Dr David and Dr John McCracken, BVMS, both of whom encouraged and inspired me to follow my

veterinary research dreams - I hope they would be proud. I dedicate this thesis to my parents, Christopher and Edith – thank you for supporting every single one of my endeavours to date. I would not be where I am today without your love and support. Love and thanks also to my siblings, Alasdair and Gillian, for always ‘keeping it real’.

And finally, with love for Andrew, without whose constant support, encouragement and enduring belief in my ability, this would have been a very different journey – thank you.

Contents

List of Figures

List of Tables

Abbreviations

AAT	African animal trypanosomiasis
AEZ	agroecological zone
bp	base pairs
BLAST	Basic Local Alignment Search Tool
CCS	circular consensus sequence
cDNA	complementary deoxyribonucleic acid
CI	Confidence Interval
CO_2	Carbon Dioxide
ConA	Concanavalin A
CLR	continuous long read
DDT	Dichlorodiphenyltrichloroethane
DNA	Deoxyribonucleic acid
dNTPs	Deoxyribonucleotide Triphosphate
dsDNA	double-stranded deoxyribonucleic acid
ECF	East Coast fever
ELISA	Enzyme-linked immunosorbent assay
FCS	Foetal Calf Serum
FMD	Foot and mouth disease
GR	Game Reserve
GIS	Geographic information system
GPS	Global positioning system
IFA	Indirect fluorescent antibody
ILRI	International Livestock Research Institute
INDEL	insertion or deletion
ITM	Infection and Treatment Method
kb	kilobase
km	kilometre
LB	Luria broth
LGCA	Loliondo Game Controlled Area
LRT	Likelihood ratio test

LSTM	Liverpool School of Tropical Medicine
MHC	major histocompatibility complex
MOI	Multiplicity of infection
ng	nanogram
NCA	Ngorongoro Conservation Area
NCBI	National Centre for Biotechnology Information
NEB	New England Biolabs®
nPCR	Nested Polymerase chain reaction
OR	Odds ratio
PCA	Principal Component Analysis
PCR	Polymerase chain reaction
PIM	polymorphic immunodominant molecule
QC	Quality control
RE	Restriction enzyme
RFLP	Restriction fragment length polymorphism
RLB	Reverse line blot
SD	standard deviation
SENAPA	Serengeti National Park
SGS	Second-generation sequencing
SMRT	Single-molecule real-time
SMS	Single-molecule sequencing
SNP	single nucleotide polymorphism
SOC	Super Optimal broth with Catabolite repression
TANAPA	Tanzania National Parks
TAWIRI	Tanzania Wildlife Research Institute
TBD	Tick-borne disease
TE	Tris-EDTA
TGS	Third-generation sequencing
TNP	Tarangire National Park
TpM	*T. parva* Muguga
TPRI	Tropical Pesticides Research Institute
TZS	Tanzanian Shilling

VPRG	Veterinary Parasite Resistance Group
ZMW	zero-mode waveguide
°C	Degrees Celsius
g	gravitational force
µg	microgram
µl	microliter
µM	micromolar
ml	millilitre
%	percentage
pmol	picomole

Chapter 1: General Introduction

1.1 East Coast Fever

East Coast fever (ECF) is caused by *Theileria parva*, a tick-borne protozoan parasite belonging to the Apicomplexa phylum (Morrison and McKeever 2006). *Rhipicephalus appendiculatus,* the three-host tick, transmits *T. parva* to infect cattle and African buffalo (*Syncerus caffer*) in Eastern, Central and Southern parts of Africa (McKeever and Morrison 1990).

The African buffalo is considered the natural host for *T. parva* but infected buffalo do not suffer disease, in contrast to cattle where mortality can be in excess of 90% (Brocklesby 1961). Buffalo are indigenous bovids of sub-Saharan Africa and have therefore co-existed with both *Rhipicephalus appendiculatus* and *T. parva* long before the introduction of cattle to the area (Grootenhuis 1989), which are thought to have been first domesticated in Africa around 10,000 years ago (Freeman et al. 2006). Buffalo living in endemic areas are essentially all infected with *T. parva* (Young et al. 1978). When buffalo-derived *T. parva* (previously referred to as *T. parva lawrencei*) is transmitted to cattle, they rapidly develop clinical disease (often referred to as Corridor disease), but often the parasites do not differentiate to piroplasms and therefore are not transmissible by ticks (Schreuder et al. 1977), thus there is evidence to suggest that even though there is genetic similarity between buffalo- and cattle-maintained parasites, most of the buffalo-derived parasites are unable to transmit between cattle (Pelle et al. 2011). This difference between buffalo- and cattle-derived *T. parva* is thought to be due to the long period of evolutionary adaptation by buffalo, compared to the relatively recent introduction of cattle into Africa (Gifford-Gonzalez and Hanotte 2011). The African buffalo (*Syncerus caffer*) is host to many diverse species and strains of *Theileria* (Allsopp and Allsopp 1988, Conrad et al. 1987, Hemmink 2014).

1.2 Life-cycle of *T. parva*

T. parva is transmitted by the tick vector, *Rhipicephalus appendiculatus,* via a complex lifecycle (Figure 1.1). Ticks become infected by feeding on infected cattle and ingesting blood containing erythrocytes infected with the piroplasm stage of the parasite (Young et al. 1986). Ticks can become infected during the larval and nymphal stages, which then transmit infection as nymphs and adults respectively (Norval et al. 1992).

Infected *R. appendiculatus* transmit *T. parva* infective sporozoites via saliva, while feeding. After several days of feeding, non-motile sporozoites are released and infect host lymphocytes, in which they rapidly escape from the endocytic vesicle into the host cytoplasm where they develop over several days into the multinucleate schizont stage (Fawcett et al. 1984).

Schizonts stimulate proliferation of the infected host cells and by associating with host microtubules are able to divide synchronously with the host cell. This process results in rapid proliferation of the infected lymphocytes (Dobbelaere and Kuenzi 2004).

Infected lymphocytes accumulate in the lymph node draining the site of infection and disseminate to both lymphoid and non-lymphoid organs via the vascular circulation system. Approximately 12-14 days after infection, schizonts develop into merozoites, which are released from infected cells, and invade red blood cells where they develop to the tick infective piroplasm stage. Piroplasms are ingested by feeding ticks and the gamete forms of the parasites undergo syngamy to form a diploid zygote, which then becomes a kinete that infects the salivary glands of the tick. At the next tick moult (stage) and feed, sporogony takes place in the salivary gland acinus, producing thousands of sporozoites, and it is during this process of sporogony that two meiotic divisions take place to yield haploid progeny. Sexual recombination therefore occurs during sporogony. After approximately four days of feeding, infective sporozoites are released into the animal (McKeever 2006, McKeever 2009). One significance of this ability of *T. parva* to recombine is its contribution to generation of antigenic diversity (Morrison et al. 2015). Due to the common

occurrence of mixed infections in the field, feeding ticks are likely to ingest a mixture of parasites (Muleya et al. 2012) allowing genetic exchange during sexual recombination (Katzer et al. 2006).

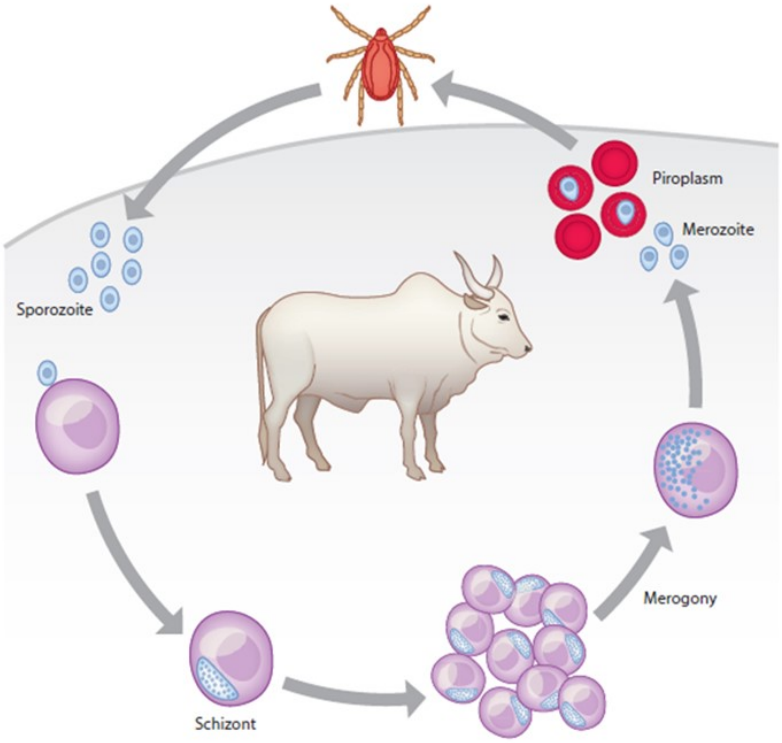

Figure 1.1: Lifecycle of *Theileria parva*.
Figure from Morrison (2015b) with permission from the Annual Review of Animal Biosciences

1.3 The tick host

Tick vectors in Tanzania have been studied for decades; however, much of the knowledge on the distribution of Tanzania's most economically significant ticks was amassed in the 1950-1980s with relatively few recent studies (McCulloch et al. 1968, Tatchell 1986, Yeoman 1966a, Yeoman 1966b, Yeoman and Walker 1967). In Tanzania, *Rhipicephalus* and *Amblyomma* genera are considered the most important in the spread of tick-borne diseases (TBDs) and are present in almost all livestock areas (Lynen et al. 2007, Lynen

et al. 2008). As well as directly transmitting diseases, ticks can cause localised necrosis to hides, reducing their quality (Jongejan and Uilenberg 2004), reduced weight gain (Pegram and Oosterwijk 1990) and milk production (Sajid et al. 2007) as well as secondary bacterial infections (Ambrose et al. 1999).

The main vector of *T. parva* is *R. appendiculatus*, a three-host tick. *R. appendiculatu*s feeds on three different hosts, detaching and moulting into the next life stage on vegetation (larva, nymph, adult) (Norval et al. 1992). It is commonly referred to as the brown ear tick because of the predilection of adult ticks to feed on the ears of cattle (Walker 2003). *T. parva* transmission occurs trans-stadially i.e. larva to nymph or nymph to adult (Blouin and Stoltsz 1989). High levels of infection with *T. parva* is detrimental to the fitness of *R. appendiculatus*, with infected ticks producing fewer eggs, and feeding and moulting happening more slowly (Watt and Walker 2000). Infection has also been shown to adversely affect the normal questing behaviour (Busby et al. 2012). Infected ticks mount an immune response against the parasite and only a very small proportion of the ingested parasites successfully infect the salivary gland (Watt and Walker 2000); this stage in the life cycle is thought to create a substantial bottle neck for the *T. parva* population in the tick vector (McKeever 2006, Pelle et al. 2011).

R. appendiculatus is found in temperate and humid habitats, from sea level to approximately 2,300 metres (FAO 1984). It inhabits eastern Africa, from southern Sudan to South Africa, extending into central Democratic Republic of Congo (DRC) and Zambia (Walker 2003). Areas in the south of the distribution, with distinct monomodal rainfall patterns, tend to have highly seasonal tick population dynamics, whereby adults are commonly only observed in the rainy season, followed by larvae and nymphs appearing in the dry months. Rainfall is the main stimulus for activity, with the seasonal pattern dictated by unfed adult ticks commencing behavioural diapause and only conducting host-seeking once the rains begin (Walker 2003). In more equatorial regions of the distribution, the rainfall pattern can be more extended

and so adults can be seen throughout the year, with no clear seasonal pattern (FAO 1984, Norval et al. 1992, Walker 2003).

Closely related to *R. appendiculatus* is *Rhipicephalus zambeziensis*, another tick capable of transmitting *T. parva*. Although mainly present in south eastern Africa, it is also present in central Tanzania where the hot and dry climatic conditions are more suitable for this species (Walker 2003).

In addition to the African buffalo, *Rhipicephalus appendiculatus* has other wildlife hosts; eland, waterbuck, greater kudu, sable antelope and nyala have been described as the main alternative hosts (Walker 2003). Waterbuck are a particularly common host for *R. appendiculatus* (Norval et al. 1992) and it has been confirmed experimentally that *T. parva* can infect waterbuck (Stagg et al. 1994, Stagg et al. 1983). Recent studies by Githaka et al. (2014) in Kenya demonstrated that although 96% of waterbuck samples (n = 26) were positive for *Theileria* species there was no detection of *T. parva* using species-specific DNA probes, suggesting that waterbuck do not play a major role in the transmission of *T. parva.* However, any role that non-buffalo wild hosts play in the epidemiology of *T. parva* requires further study, as current evidence is limited.

1.4 Pathogenesis and clinical signs

The pathogenesis of *T. parva* infection is caused by schizont-infected lymphocytes (Norval et al. 1992). Schizogony of *T. parva* takes place in B- and T-cells (Baldwin et al. 1988), but pathogenesis is caused predominantly by infected T cells (Morrison et al. 1996a). Rapid multiplication of the infected cells is dependent on live parasites, as anti-theilerial treatment halts the proliferative state (Brown et al. 1989a, Dobbelaere et al. 1991). Progressive parasite multiplication results in extensive lymphocytolysis and leukopenia (Morrison et al. 1981).

Infected lymphocytes are detectable in the lymph node draining the bite of the tick 5-10 days post-infection, depending on the dose of parasites (Irvin 1987,

Norval et al. 1992). Initial detection of infection coincides with development of pyrexia and generalised lymphadenopathy (Morrison et al. 1989). Appetite reduction or anorexia can develop in the later stages of infection. Schizont-infected lymphocytes are often detectable in blood smears approximately two weeks post-infection. Lymphoblastic infiltration of the pulmonary tissue results in profuse oedema and respiratory signs, and death usually occurs 14 to 30 days post-infection (Irvin 1987). There is often significant lymphoblastic infiltration of the gastrointestinal tract, with diarrhoea resulting. A rare form of ECF is the occurrence of neurological clinical signs, associated with the presence of schizont-infected cells and mitotic bodies in histological examination of brain tissue at post-mortem (Coetzer and Tustin 2004). Accumulation of schizonts in cerebral capillaries is thought to cause the neurological signs, so-called ‘turning sickness’ (Urquhart 1996).

In surviving cattle, the immune response produces an asymptomatic carrier state, whereby there is the persistence of low levels of infected lymphocytes, which in turn provides further sources of infected red blood cells for transmission to the tick vector (Potgieter et al. 1988).

1.5 Diagnosis in mammalian hosts

Theileria species were traditionally described on the basis of microscopic examination of Giemsa stained blood smears (Mans et al. 2015). Microscopy, however, is unable to detect carrier animals due to their low piroplasm parasitaemias (Norval et al. 1992, Zweygarth et al. 1997) and is also unreliable for speciation due to the morphological similarities of piroplasms (Lawrence 1979, Uilenberg 1981). In the 1990s the indirect fluorescent antibody (IFA) test was commonly used to detect schizont antigen, but this test was not able to differentiate between several different *Theileria* species (Goddeeris et al. 1982, Grootenhuis et al. 1979). Similar to microscopy, serological tests, including the IFA, are not always able to detect carriers (Dolan 1986). The definitive method for detecting carrier state is by xeno-diagnosis, the

application of ticks to naïve mammalian hosts and confirmation of the subsequent transmission of infection (Bishop et al. 1992).

An antibody enzyme-linked immunosorbent assay (ELISA) was developed to identify a *T. parva* schizont surface protein (polymorphic immunodominant molecule, PIM) that induces a strong antibody response (Katende et al. 1998). The ELISA test identifies exposed individuals but not necessarily individuals with current infection. A reverse line blot (RLB) PCR assay amplifies the 18S rRNA gene, detecting all *Theileria* and *Babesia* species in cattle by species-specific probe hybridisation and chemiluminescence (Gubbels et al. 1999, Oura et al. 2004b). Real-time PCR, amplifying the 18S rRNA gene, is used for detecting *T. parva* infections in cattle and buffalo (Sibeko et al. 2008) and most recently a nested PCR assay, amplifying the p104 gene, was developed (Odongo et al. 2010, Skilton et al. 2002). The DNA-based tests have a theoretical advantage over serology-based tests in that they detect the presence of parasite DNA (i.e. are markers of active infection), but can suffer from sensitivity issues due to the low number of parasites, particularly in carrier animals.

Diagnostic assays must be reliable and have evidence of effectiveness in order to provide robust and accurate data (Bustin 2010, Peeling et al. 2010) and these will be discussed in detail in Chapter 3.

1.6 *Theileria parva* carrier state

Theileria parva is endemic in many areas of eastern Africa (Coetzer and Tustin 2004). The term ‘endemic’ refers to the constant presence, and often predictable level, of disease (Thrusfield 2005). In areas where there is a constantly high level of infection in all age-groups, the term ‘hyperendemic’ is applied (Thrusfield 2005).

Endemic stability is described as an epidemiological state where there is little clinical disease despite the majority of the population being infected (Norval et al. 1992). It can be described as a balance “where host, disease, vector and

environment coexist resulting in minimal clinical disease" (Norval et al. 1992). In the case of *T. parva*, endemic stability depicts the epidemiological state in a population where the majority is exposed to infection by six months of age, and develop immunity, resulting in little or no clinical disease (Norval et al. 1992). In contrast, endemic instability is the state when exposure to infection is low and clinical disease is common (Norval et al. 1992).

In the case of ECF, a carrier state exists in animals that recover from the initial infection and clinical disease. Despite recovering from infection, there remain a low number of infected lymphocytes present within lymphoid tissues (Morrison 2007, Skilton et al. 2002) and there is persistence of this population of infected lymphocytes, which in turn give rise to low levels of tick-transmissible piroplasms. Young et al. (1981, 1986) first showed that recovered cattle could transmit fatal ECF via ticks to naïve cattle in an area free from buffalo, and Kariuki et al. (1995) demonstrated the presence of persistent schizonts and piroplasms in immunised cattle after infection with *T. parva*, these studies dispelling the previous belief that recovered animals were not a source of infection (Uilenberg 1984). Ticks feeding on carrier animals ingest only a few erythrocytes and so such ticks also show low levels of infection (Norval et al. 1992).

It has been known for many years that there similarly exists a carrier state of *T. parva* in buffalo (Barnett and Brocklesby 1966, Young et al. 1978). The carrier state is considered to be of great importance in the maintenance of *T. parva* in the field since in endemic areas most tick feeds are on carrier animals rather than acute infections (Young et al. 1986).

1.7 Treatment and control

There are limited treatment options for East Coast fever and those that exist are expensive and limited by the necessity to be used in the acute stage of the disease. Once the disease is too advanced, the animal is unable to recover following therapy because of extensive destruction of the immune system.

Theilericidal compounds such as parvaquone or buparvaquone can be used successfully if administered in the early clinical stages of the disease (Muraguri et al. 1999). Availability and cost of chemotherapeutic drugs is also often prohibitive to smallholder farmers (Norval et al. 1992). As well as the licensed theilericides parvaquone or buparvaquone, oxytetracycline can be an effective treatment if administered immediately following infection (Dolan et al. 1984).

Oxytetracycline is used as part of the 'infection and treatment method' (ITM) of vaccination with live parasites, which involves inoculation with live sporozoites while simultaneously being treated with long acting oxytetracycline (Radley 1975a, Radley 1975b, Radley 1975c). This vaccination provides long-lasting immunity against homologous *T. parva* isolates (Burridge et al. 1972, Morzaria et al. 1987b) but not all immunised animals are protected against heterologous isolate challenge (Radley 1975a). Based on a series of studies involving immunisation and cross-challenge with different parasite isolate combinations, Radley et al (1975c) identified a mixture of three isolates, which is referred to as the Muguga cocktail, that provided broad protection to cattle challenged with heterologous *T. parva* isolates and field challenge. However, protection is not conferred to all strains and is often not conferred to buffalo-derived strains (Bishop 2015, Radley 1981, Radley et al. 1979, Sitt et al. 2015). More recent studies by Norling et al (2015) and Hemmink et al (2016) have identified remarkable similarities between two of the three Muguga cocktail stocks – Serengeti-transformed and Muguga, with sequenced genomes showing only 420 non-synonymous single nucleotide polymorphisms (SNPs) – while the third isolate, Kiambu5, is significantly more divergent, with almost 40,000 SNPs relative to the reference Muguga genome. Despite the ability of the vaccine to confer protection there are, however, practical constraints in the use of the vaccine, including the necessity for a cold chain, which have limited its use (Morrison 2015, Morrison et al. 2015). An additional concern is that the vaccine is capable of establishing a carrier state and therefore may introduce the vaccine parasite strains into local tick populations (Di Giulio et al. 2009, McKeever et al. 1999). Further reference will be made to the vaccine in later chapters.

The most commonly used method of controlling TBDs in the tropics is acaricide treatment (De Castro 1997). There are several classes of acaricides licensed and available, namely organophosphates and carbamates, pyrethroids and amidines (De Meneghi et al. 2016). Acaricides are used to control tick infestation but this requires regular acaricide application, making it an expensive and a practically challenging method to sustain. The frequent application of acaricide is also conducive to the development of acaricide resistance (Jongejan and Uilenberg 2004), as well as raising concerns about acaricide residues in the animal products and environmental contamination (Mugisha et al. 2005). The most commonly used application methods for acaricides are by treating animals with dip or spray (FAO 1984).

In Tanzania, the Government has adopted a policy that enforces tick control in attempts to prevent ticks being disseminated via cattle movement and exportation (MLFD 2015). Use of dip tanks was reported as early as 1905, with over 2,000 dip tanks constructed across Tanzania by 1990; however subsequent lack of tank management and maintenance resulted in only 5% of tanks being functional by 1991 (Lynen et al. 2007).

When acaricides are being used intensively, a dependence on tick prevention develops as cattle are highly susceptible to TBDs if they are not regularly exposed to ticks. Therefore, any subsequent breakdown in tick control can result in significant outbreaks of disease in the cattle population (Jongejan and Uilenberg 2004).

1.8 Impact

T. parva is considered the most important of the *Theileria* species that infects cattle in sub-Saharan Africa (Coetzer et al. 1994) and it is the most economically important disease of cattle in this region (Norval et al. 1992); this is in part due to it being a major hindrance to the introduction of exotic cattle breeds due to their higher level of susceptibility to ECF (Minjauw 2003). Approximately one million cattle are lost from ECF annually (Mukhebi et al.

1992). The impacts of ECF losses are greatest to small-scale farmers and the annual cost of mortality, chemotherapeutic treatment and control is estimated to be in the region of US $300 million (Minjauw 2003).

Most deaths due to ECF occur in calves, of less than a year old (Barnett 1957, Gitau et al. 1999, Moll et al. 1986, Swai et al. 2009). A longitudinal study of mixed breed dairy cattle by Swai et al. (2009) in eastern Tanzania observed a 12% mortality rate at one year of age and found 56% was due to TBDs, with 38% of those attributed to ECF. They also reported 8% of calves (44 of 549) experiencing at least one case of clinical disease during the first year, with 43% of those (19 of 44) experiencing ECF (Swai et al. 2009). The Infectious Diseases of East African Livestock (IDEAL) project followed a cohort of calves in western Kenya over a one year period and observed a 16% mortality rate (88 of 548), with 36% of those due to ECF (32 of 88) (de Clare Bronsvoort et al. 2013). Due to the losses associated with the morbidity and mortality caused, ECF is a significant burden to the small holder farmer.

Pastoralists in east Africa are highly dependent on livestock for income and they place high value on ownership of livestock. These farmers are often vulnerable and marginalised with land use constraints and regular environmental challenges such as droughts (Minjauw 2003). Tick-borne diseases such as *T. parva* can cause particular problems when an unstable disease situation develops, resulting in a high number of animal deaths; this tends to impact most on the poorest farmers who may not be able to afford veterinary attention and treatment costs. When draught animals are ill, crop production is also affected and so income losses from both livestock and crop production will ensue (Minjauw 2003).

An additional challenge that farmers face in the region is the diversity of parasites that cattle are exposed to, frequently resulting in co-infections with multiple pathogens (Petney and Andrews 1998). For example, co-infection with trypanosomes has been shown to cause a six-fold increase in the risk of death by ECF (de Clare Bronsvoort et al. 2013, Thumbi et al. 2014, Woolhouse et al. 2015).

1.9 Other *Theileria* species

Although *T. parva* is considered the most important of the *Theileria* species, there are several other species found in buffalo, which can infect cattle (Reviewed in Morrison 2015a). *Theileria taurotragi*, originally isolated from eland and also found in buffalo, is transmitted to cattle principally by *R. appendiculatus* and is usually benign (Coetzer and Tustin 2004). *Theileria mutans* and *Theileria velifera,* both transmitted predominantly by *Amblyomma hebraeum,* are widespread in cattle and buffalo, and are generally considered to be non-pathogenic, although there have been reports of pathogenic strains of *T. mutans* in cattle (Snodgrass et al. 1972). *Theileria buffeli,* transmitted by *Haemaphysalis* ticks, infects buffalo and cattle. *Theileria* sp. (buffalo), which is closely related to *T. parva*, is found in buffalo (Allsopp et al. 1993, Ngumi et al. 1994, Oura et al. 2011b) and recently has been described also in cattle (Bishop 2015, Hemmink 2014). In a study by Oura et al. (2011b) in Uganda it was observed that all buffalo examined were infected with *T*. sp. (buffalo) whereas it was not detected in any of the cattle. However Hemmink (2014) and Bishop et al. (2015) demonstrated the presence of *T*. sp. (buffalo) in cattle introduced to an area in Kenya grazed by buffalo. In these studies, since the animals were co-infected with *T*. sp. (buffalo) and *T. parva*, it was not possible to assess the pathogenicity of *T*. sp. (buffalo). Given its ability to infect and transform cattle lymphocytes, it is possible that there is associated pathology, but a pure *T*. sp. (buffalo) sporozoite stabilate would be required to test this. The tick vector of *T*. sp. (buffalo) is currently unknown but due to the frequently observed co-infection with *T. parva*, a tick vector with similar climatic conditions and geographical distribution to *R. appendiculatus* is indicated. Several have been suggested by Pienaar et al. (2011), including *Amblyomma hebraeum*, *Hyalomma truncatum*, *Rhipicephalus appendiculatus* or *Rhipicephalus simus*, but this is not an exclusive list of potential vectors.

A condition referred to as “Ormilo”, which causes neurological signs in cattle in Maasai herds in northern Tanzania, has also been attributed to infection by *Theileria* species (confusingly, “Ormilo” is also used to refer to the clinical signs

associated with *Taenia multiceps* coenurosis in small ruminants in the same area (Hughes et al. 2019)). It has been proposed that *T. taurotragi* may be the cause of this neurological disease but further investigation is required to definitively identify the causative parasite (Catalano et al. 2015).

1.10 Immunity to *T. parva*

T. parva sporozoites rapidly invade bovine lymphocytes by receptor-mediated endocytosis and then escape from the endocytic vacuole to enter the cytoplasm where they develop to schizonts (Fawcett et al. 1982). Schizonts stimulate proliferation of the host cells, during which they divide synchronously with the host cells, allowing them to remain in an intracellular location throughout this stage of development. Hence, sporozoites are exposed to serum antibodies for only a brief window of time and the schizont stage is not exposed to the extracellular environment of the host and so is “hidden” from a humoral response. Antibody is, therefore, not considered to have a major role in the immune response to *T. parva* (McKeever et al. 1999, Muhammed et al. 1975). Neither maternal transfer of antibodies nor experimental transfer of immune serum provide immune protection (Muhammed et al. 1975). By contrast, there is strong evidence that immunity involves T cell-mediated immune responses (McKeever et al. 1999).

It was shown that cattle can be immunised against *T. parva* by infection and simultaneous treatment with a slow release formulation of oxytetracycline. Animals immunised with a single parasite isolate developed immunity to homologous parasite infection. However, incomplete protection was seen when challenged with heterologous isolates, with some animals immune and some susceptible (Radley 1975a).

Studies of the mechanisms of immunity to *T. parva* have indicated that immunity is targeted at the intra-lymphocytic schizont-stage (Morrison and McKeever 1998), since schizont-infected cells are detected around 12-16 days after immunisation by infection and treatment, prior to control of the infection

(Eugui and Emery 1981). Strong parasite-specific T cell responses are detected at the time of parasite clearance.

Infected lymphocytes express high levels of class I and class II MHC molecules on the cell surface and elicit parasite-specific CD4 and CD8 T cell responses. CD8 T cells, restricted by class I MHC, show a strong MHC-restricted cytotoxic activity (Goddeeris et al. 1986a, Goddeeris et al. 1986b, Morrison et al. 1987).

Emery (1981) demonstrated that transfer of thoracic duct lymphocytes from immune to naïve twin calves conferred immunity against challenge with *T. parva*. Moreover, McKeever et al. (1994) showed that adoptive transfer of CD8 T cell-enriched populations between immune and naïve twin calves also resulted in protection, demonstrating the protective activity of the *T. parva*-specific CD8 T cells. Although there is no direct evidence for CD4 T cell responses in immunity, *in vitro* studies have indicated that they are required for effective activation of specific memory CD8 T cells (Taracha et al. 1997).

Parasite strain-restriction of CD8 T cell specificity has been shown in experiments where animals were immunised with one parasite strain and subsequently challenged with a different strain; some animals were protected but others were not (Morrison et al. 1987, Radley 1975a). A correlation was found between the strain specificity of the CD8 T cell response in these animals and their susceptibility to challenge with the heterologous parasite (Taracha et al. 1995a, Taracha et al. 1995b); individuals with cross-reactive responses showed protection to challenge compared to those with strain-specific CD8 T cells who showed clinical reactions. This finding suggested that the antigens recognised by the detectable CD8 T cell response are significant in generating immune protection (Morrison 2007).

Detailed studies of CD8 T cell responses in immune animals indicate that the restricted specificity of the response is due to the response of the individual animal being focused on a few highly dominant antigens. These antigens display genotypic polymorphism, and depending on the individual's MHC type, different antigens are recognised by the CD8 response (Goddeeris et al. 1990,

Morrison et al. 1996a, Taracha et al. 1992) and hence responses of animals of different MHC types show varying patterns of strain specificity (Goddeeris et al. 1990, Taracha et al. 1995b). This immunodominance is a commonly reported feature of CD8 T cell responses to viruses in humans and mice (Yewdell 2006).

1.11 *T. parva* antigens

Studies have identified *T. parva* antigens that are recognised by CD8 T cells from immune cattle (Graham et al. 2006), which has facilitated investigation into the immune response to *T. parva* (Morrison et al. 2015). Antigens were identified by two approaches; the first was a targeted gene approach, whereby genes predicted to contain a secretory peptide signal (using preliminary genome sequence data) were screened by detection by CD8 T cells from immune animals. This was based on the assertion that proteins secreted by *T. parva* schizonts into the cell cytoplasm are likely to access the MHC class I antigen-presenting pathway of the host cell. The second approach was a high-throughput screen of a schizont cDNA library co-expressed with relevant class I MHC genes in Cos7 cells. These approaches identified eight CD8 T cell target antigens (Tp1-Tp8) (Graham et al. 2006). During the screening studies, which utilised *Bos taurus* and *Bos indicus* cattle T cells, it was observed that there was differential detection of antigens by animals with different MHC types (Graham et al. 2008). Similarly, when the antigens were screened for epitopes, only one or occasionally two epitopes recognised by CD8 T cells were identified for each MHC. When T cell lines derived from immune Holstein cattle were screened, another three antigens were identified (Tp9, Tp10 and Tp12) with a similar pattern of variable MHC type recognition (Hemmink et al. 2016) (unpublished MacHugh, N., Graham, S. and Morrison, W.).

The identification of *T. parva* antigens and epitopes important in effective immune responses has enabled further studies on the immune response and associated selective pressures that stimulate antigenic diversity, and have allowed investigation of allelic polymorphism in antigen-encoding genes.

1.12 Antigenic diversity of *T. parva*

Early studies of *T. parva* with parasite-specific monoclonal antibodies and DNA probes had indicated that the parasite displays genetic and antigenic diversity (Bishop et al. 1993, Bishop et al. 1994a, Geysen et al. 1999). However, these methods inevitably focused on small parts of the genome. The subsequent identification of satellite DNA markers across the parasite genome provided more robust tools to examine genetic diversity (Oura et al. 2005, Oura et al. 2003) and then the identification of antigens recognised by immune CD8 T cells provided a means of examining immunologically relevant genetic diversity in *T. parva* (Graham et al. 2007, Graham et al. 2006). Ultimately, sequencing of *T. parva* genomes has allowed detailed genomic comparisons between parasite strains (Gardner et al. 2005, Hayashida et al. 2013).

The heterogeneity of *T. parva* genotypes has been observed in cattle studies from a range of east African countries (Bishop et al. 1993, Geysen et al. 1999). Later studies with mini- and micro-satellite markers reported by Odongo et al. (2006) identified a high level of *T. parva* genetic diversity within cattle in Kenya, but with minimal sub-structuring in relation to geographical origin. Oura et al. (2005) also identified high levels of parasite diversity in three separate regions of Uganda, with evidence of geographical sub-structuring between two regions as well as sub-structuring within one region.

Molecular characterisation of buffalo-derived *T. parva* has shown high levels of genotypic diversity. Studies by Bishop et al. (1994a), Collins and Allsopp (1999), Geysen et al. (2004) and Oura et al. (2004a) have all demonstrated greater diversity in buffalo-derived *T. parva* parasites than in cattle-derived. Oura et al. (2011b) observed greater *T. parva* diversity in the buffalo samples than the cattle and demonstrated that the buffalo-derived *T. parva* population was separate from the cattle-derived parasite population in Uganda, with no shared genotypes.

MacHugh et al. (2009) demonstrated that the Tp1 and Tp2 antigens are highly dominant targets for CD8 T cells in cattle of MHC types A18 and A10, respectively, and extensive genetic diversity has been observed in these two

antigens. By Sanger sequencing of PCR amplicons of the same genes encoding Tp1 and Tp2 from a range of parasite isolates of both cattle and buffalo origin, Pelle et al. (2011) detected multiple allelic variants of these two *T. parva* antigens. The majority of alleles were found in isolates from buffalo, or cattle that had developed acute infections while grazing alongside buffalo, and there was less allelic variation in isolates from cattle with no buffalo association. Nucleotide sequence diversity also resulted in extensive diversity in the amino acid sequences, including the known CD8 T cell epitopes identified within Tp1 and Tp2 (Pelle et al. 2011).

A subsequent study by Hemmink (2014) examined diversity in Tp1 and Tp2 and a further four CD8 T cell antigen genes (Tp4, Tp5, Tp6 and Tp10) in samples from Kenyan and South African buffalo, using high-throughput sequencing of PCR amplicons. The results revealed extensive genetic diversity of *T. parva* in both the Kenyan and South African buffalo, both at the population level as well as within individuals. As with the results reported by Pelle et al. (2011), Tp1 and Tp2 were found to have a high degree of antigenic diversity (76 alleles of Tp1 and 70 alleles of Tp2), with many epitope variants found in Tp2 (31, 34 and 42 variants of the three epitopes, respectively) and three variants found in Tp1. By contrast, the extent of sequence diversity in the other four antigens was more limited. However, multiple alleles of these genes were detected, although they showed much more limited diversity at the amino acid level (Hemmink et al. 2018). Hence, unlike Tp1 and Tp2, no or limited antigenic diversity was found in these four antigens.

The observed differences in levels of *T. parva* diversity in cattle and buffalo has important implications with regards to strain specificity of immunity. Early experimental studies revealed incomplete protection of cattle immunised with cattle-derived *T. parva* that were subsequently challenged with buffalo-derived parasites (Young et al. 1973). More recently, studies by Bishop et al. (2015) and Sitt et al. (2015), who introduced cattle vaccinated with the Muguga cocktail into areas grazed by buffalo, found that the cattle were poorly protected against the disease, thus implying that the buffalo-derived parasites

that the cattle were exposed to were antigenically distinct. Nevertheless, the Muguga cocktail vaccine has been used successfully to control ECF in parts of Northern Tanzania (Di Giulio et al. 2009, Martins et al. 2010) where buffalo are present although it remains uncertain as to the extent to which these cattle were challenged with buffalo-derived *T. parva*.

1.13 Disease control at the livestock/wildlife interface

Infectious diseases are transmitted and spread between wildlife and domestic animals via direct contact and indirectly via arthropod vectors such as ticks, tsetse flies and mosquitoes (Kock 2005). In addition to being the natural host to *T. parva*, the African buffalo is host to several other economically important livestock diseases at the wildlife/livestock interface, prime examples being foot and mouth disease (FMD), bovine tuberculosis (*Mycobacterium bovis*) and bovine brucellosis (*Brucella abortus*) (Norval et al. 1992). Indigenous pathogens (i.e. those that originated in Africa), have co-evolved with wildlife hosts for a long time, and therefore the interaction between host and pathogen rarely results in severe clinical disease and threat to the host's survival; buffalo-maintained indigenous diseases include East Coast fever, foot and mouth disease (FMD) and animal African trypanosomiasis (AAT) (De Vos and Bengis 1994, Thomson 1995). In contrast, diseases such as bovine tuberculosis and bovine brucellosis have been introduced via livestock, and have established in buffalo which then act as a wildlife reservoir (De Vos et al. 2001). Buffalo are a social species and often reside in large herds of up to 1000 animals (Michel et al. 2006). However, smaller groups are often apparent and are known to migrate over large distances in search of water and as part of the dispersal strategy of heifers or bachelor bulls, and this movement of buffalo adds to the dispersal of pathogens (Cross et al. 2005).

Disease control at the livestock/wildlife interface is extremely challenging, especially with increasing competition for land available for livestock (Michel and Bengis 2012). In Botswana and southern Africa there have been disease control efforts to separate wildlife and cattle by use of 'veterinary fencing'

(Mogotsi et al. 2016). FMD outbreaks in southern Africa are due to infection from buffalo, where the virus persists in the buffalo without clinical disease (Hedger 1972) and the diversity of virus serotypes found in buffalo, when combined with unrestricted animal movement, result in a disease that is difficult to control (Vosloo et al. 2002). In Tanzania, however, FMD is mostly driven by circulation within livestock populations, due in part to cattle movements (Casey-Bryars et al. 2018). To date there has been little research into the role of wildlife or cattle host migration in the transmission of tick borne diseases (Wamuyu et al. 2015).

In buffalo/livestock interface areas there are limited ECF control methods available. Veterinary fencing is used commonly in southern Africa to control infectious disease transmission between wildlife and livestock, but this approach has not been taken up in eastern Africa. The avoidance of buffalo areas is likely to reduce contact with *T. parva* infected ticks but this can restrict grazing availability in shared areas. Effective tick control practices in local livestock populations will provide protection to cattle, and immunisation with the ITM vaccine can also confer protection, although the vaccine efficacy in buffalo areas is uncertain (Nene and Morrison 2016). Specifically, there are a lack of data collected on different epidemiological and genetic parameters within an ecosystem that enable a comprehensive understanding of ECF epidemiology and control in livestock/wildlife interface areas. It is a real challenge when there must be compatibility between livestock farming as well as wildlife conservation at livestock/wildlife interfaces (Michel and Bengis 2012).

1.14 Key questions regarding ECF control at the livestock/wildlife interface

Improving control of ECF requires an improved understanding about the presence and diversity of *T. parva* in both cattle and buffalo, and the role that is played by the interface where both hosts, and the tick vector, can interact. This is particularly true at the moment; there is a general trend of increased

encroachment of livestock into protected areas, hereinafter referred to as “wildlife areas”, allowing for increased interaction between hosts and pathogen transmission and presenting increased challenges for disease control (Rhyan and Spraker 2010).

In this study, the livestock/wildlife interface is investigated for the prevalence of *T. parva* in both cattle and buffalo, as well as examining the diversity of the *T. parva* population at this interface. The study uses samples from two areas where cattle and buffalo interact; Serengeti National Park, Tanzania (described fully in Chapter 2) and Ol Pejeta game conservancy, Kenya (described fully in Chapter 3).

There have been studies carried out at Ol Pejeta previously, specifically investigating the transmission of *T. parva* from buffalo to cattle (Sitt et al. 2019, Sitt et al. 2015). The Ol Pejeta samples used in this study were collected in 2014 as part of a vaccine trial (unpublished, personal communication Phil Toye, ILRI, Kenya) and the samples were stored and extracted in the same manner as described by Sitt et al. (2015). This study site presents an unusual opportunity to study the effects of buffalo-derived pathogens, since it had only been grazed by buffalo prior to the introduction of cattle and so *T. parva* infections observed in cattle must be derived from buffalo. The epidemiology of *T. parva* at the interface of the Serengeti National Park, however, has not previously been studied and in order to establish effective control measures, there is the requirement for awareness and understanding of the disease in this area.

1.15 Aims and objectives

The overall aims of the project are threefold, first to ascertain the prevalence of *T. parva* in cattle and buffalo in an area in north-east Tanzania, on the border of the Serengeti National Park, and second, to determine which vector and disease control practices are being used in this area. Third, as the study area is host to both cattle and buffalo, it provides an ideal opportunity to examine

genetic and antigenic diversity of field populations of *T. parva* at the livestock/wildlife interface.

The specific objectives of the project were to:

1. Determine the prevalence of *T. parva* in cattle and buffalo in and around the Serengeti Ecosystem, and assess risk factors associated with the presence of *T. parva* infection (Chapter 2).
2. Identify vector control practices being used in the study area and look at herd-level risk factors influencing uptake of control (Chapter 2).
3. Design a pipeline for amplifying and analysing full-length or near full-length target antigen sequences (Chapter 3)
    a. Establish a robust diagnostic test to detect *T. parva* in field samples, using samples from Ol Pejeta game conservancy
    b. Design primers to amplify antigen genes across diverse *T. parva* strains
4. Apply the genotyping pipeline to samples from cattle and buffalo in the Serengeti National Park study area to assess genetic and antigenic diversity of the *T. parva* population in this area (Chapter 4).

The collection of some of the data in the study was impossible to do alone, and therefore I was part of a team that designed and implemented the cross-sectional cattle survey. In addition to having designed the questionnaire, arranged translation and sought ethical approval, I trained the enumerator in implementing the questionnaire and was present for most of the interviews. I conducted all day to day logistics and financial arrangements for implementation of the questionnaire in the field. I was also part of a team in planning and conducting workshops. Samples collected prior to the study were provided, but I processed and exported many from Tanzania. I conducted all further sample processing, primer design and analysis. Bioinformatic processing was carried out by Siddharth Jayaraman,

bioinformatician in the L. Morrison Group, with decisions made in discussion with myself.

Chapter 2: Epidemiological Study

2.1 Introduction

ECF is a major cause of death in cattle in Northern Tanzania (Homewood et al. 1987), especially among calves in Maasai pastoral herds where mortality can be 40-80% (Di Giulio et al. 2003, Homewood et al. 2006). There are approximately 25 million cattle raised in Tanzania of which most are indigenous Tanzanian Shorthorn Zebu (Laisser et al. 2017). Cattle are farmed for a multitude of purposes, including provision of milk and meat as well as draft power and they are a vital source of income and employment for rural smallholder farmers (MLFD 2015). The losses associated with ECF can be substantial and the disease can be a significant restriction to the development of the livestock sector (Chenyambuga 2010).

Livestock production systems in Tanzania include traditional extensive livestock production referred to as pastoralism, traditional crop-livestock production referred to as agro-pastoralism, and intensive/commercial livestock production known as ranching (MLFD 2011), with pastoralism considered the most predominant livestock system (~97%) (MLFD 2015). In pastoral herds indigenous cattle are reared and as these cattle are relatively adapted to survive severe conditions, ECF prevalence can be high as minimal control measures are often taken (Ngowi 2008). Cattle in agro-pastoral systems are grazed in different sites depending on the season and ECF prevalence can be variable (Chenyambuga 2010), with increased risk of infection when moving to pasture (Laisser et al. 2014).

The epidemiology of ECF is greatly influenced by environmental conditions which in turn affect the dynamics of the tick vector (Gachohi et al. 2012). The suitability for *R. appendiculatus,* therefore, varies in different agroecological zones (AEZ) and thus tick abundance impacts the regional endemic stability of ECF (Yeoman 1966a). Yeoman (1966a) demonstrated a direct correlation between tick infestation rates and ECF epidemiological state (endemic, epidemic, sporadic and ECF-free) in Sukumaland, Tanzania. The epidemiology of *T. parva* must be understood in order to target control

measures effectively; changes in control may risk altering the balance of endemic stability.

Several studies have been carried out to establish *T. parva* prevalence in various areas of Tanzania, at differing times of year and with differing agroecological conditions. These studies are described in Table 2:1.

Table 2:1: *T. parva* prevalence studies in Tanzania, with epidemiological factors when known

Region	District	Prevalence	Detection method	Time of year	Epidemiological factors	Reference
Northern Tanzania	Tarime	16.7%	PCR		62% animals infested with ticks	(Laisser et al. 2014)
	Serengeti	38.3%	PCR		92% animals infested with ticks	
North-east Tanzania	Manyara	21.1	PCR	May (end of rainy season)	Sporadic acaricide use	(Kazungu et al. 2015a)
	Mwanza	60.1%	PCR		Regular acaricide use	
Northern Tanzania	Simanjiro	50.39%	PCR	April (end of rainy season)	ITM immunised	(Kazungu et al. 2015b)
	Simanjiro	19.69%	PCR		Non-immunised	
Northern Tanzania	Mara	21.8%	PCR	June-Sept (dry season)	*R. appendiculatus* in high numbers, not influenced by season	(Kerario et al. 2017)
Western Tanzania	Mbeya	7.4%	PCR		*R. appendiculatus* present where substantial rainfall	
Central Tanzania	Singida	13.4%	PCR		*R. appendiculatus* in high numbers during rainy season	
Eastern Tanzania	Kilosa	8.1%	ELISA	Oct-April		(Tarimo 2013)
Northern Tanzania	Tarime	80.1%	ELISA	Oct-Jan	37.5% farmers using acaricides	(Chenyambuga 2010)
North-east Tanzania	Tanga	23%	ELISA	Jan-April	57% farmers using intensive acaricide application	(Swai et al. 2007)

These studies demonstrate a wide range of prevalence, which is affected by an array of factors including agro-ecological zone (AEZ), with the associated climate, vegetation and rainfall patterns. Additionally, tick burden is affected greatly by AEZ as well as farm management practices, including tick control methods (Norval et al. 1992). The method of detecting *T. parva* can also greatly affect the prevalence, depending on an assay's sensitivity. These factors must all be taken into consideration when comparing the wide range of *T. parva* prevalence in different study sites, as well as clinical features of the sampled animals, in order to understand and determine the epidemiological state of an area.

Proximity to protected areas and wildlife can be associated with increased exposure to disease – in areas where wildlife and livestock are not separated the transmission of TBDs is highly likely (Young et al. 1985). Haji et al. (2014) observed a positive association between haemoparasites presence in cattle and proximity to protected areas. There are limited studies on ECF in livestock/wildlife interface areas, where cattle may be exposed to buffalo-derived *T. parva* in addition to *T. parva* circulating in cattle. Kazunga et al. (2015b) reported a significant association between higher numbers of *T. parva* carrier cattle and proximity to the Tarangire National Park, where livestock and wildlife are not separated. Gachohi et al. (2012) reviewed the epidemiology of ECF in Kenya and described one livestock/wildlife interface in the Trans-Mara Division, Narok District; calves were observed until six months of age, during which time 23% developed patent *T. parva* infections, but there were no deaths reported attributable to *T. parva* (Moll et al. 1984). A recent study by Nthiwa et al. (2019) investigated the diseases of concern to pastoral farmers in Maasai Mara, Kenya, comparing those reported in an area with a high degree of livestock-wildlife interaction to those diseases reported in an area with minimal interactions. ECF was reported as the second most prevalent disease in cattle (after Malignant Catarrhal Fever), although this was not found to differ between the areas. In summary, despite the risks of disease transmission from buffalo, there are a lack of data describing the epidemiology of ECF in areas where cattle and buffalo co-exist.

As well as environmental conditions, another important factor that influences the epidemiology of ECF is the vector control measures being used (Laisser et al. 2017). The main methods of control are tick control, chemotherapy and immunization with the ITM vaccine (Gachohi et al. 2012). In Tanzania the use of acaricides is commonly practiced (Kivaria 2006), however the products are often prohibitively expensive and farmers have been reported to underdose their animals to reduce costs (Laisser et al. 2014). Field studies have been carried out investigating acaricide use in Tanzania; Chenyambuga et al. (2010) reported 60% of farmers using acaricides to control ticks in a study in Kenya, Uganda and Tanzania, with Tanzanian farmers dipping or spraying their cattle when they could afford to buy acaricide. Kazunga et al. (2015a) reported acaricide use by small-holders in Manyara and Mwanza Districts in Tanzania, and Kerario et al. (2017) described acaricides being used by farmers in Mara, Mbeya and Singida regions of Tanzania, whereby cattle were being dipped and sprayed. Nonga et al. (2012) carried out a study to assess the efficacy of a cypermethrin acaricide in Mvomero district, Tanzania. A recent study by Mwaseba and Kigoda (2017) assessed tsetse control practices in villages in close proximity to the SENAPA and observed minimal use of control methods. The ITM vaccine is being used in parts of Tanzania (Di Giulio et al. 2009, Kazungu et al. 2015b, Lynen 2005).

The Serengeti Ecosystem was identified as a key area where cattle would be at risk of *T. parva* infection from buffalo in a risk mapping exercise based on the distribution of cattle, buffalo, *R. appendiculatus* and *T. parva*. (Wint and Kiara 2017) (Figure 2.1).

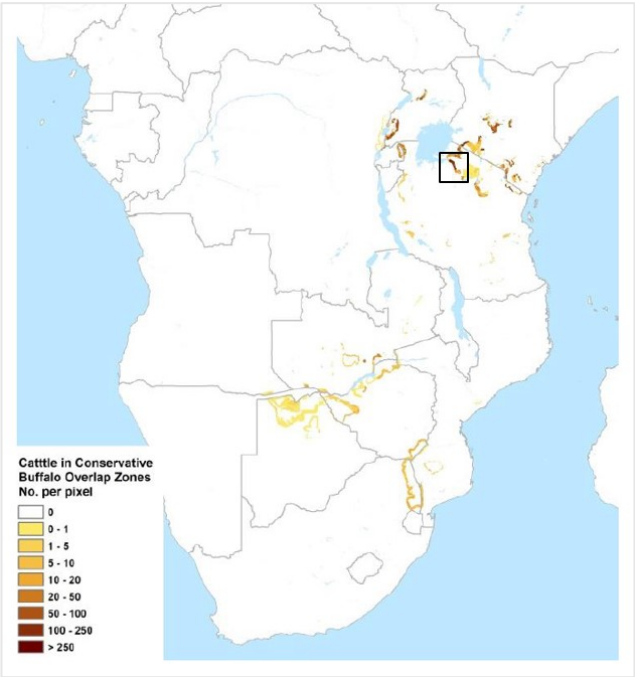

Figure 2.1: Map of southern Africa showing cattle in buffalo-derived *T. parva* risk zones.
Study area is located within the black outlined box. Map adapted with permission from Wint and Kiara (2017).

There are no published studies describing the prevalence of *T. parva* at the boundary of the Serengeti National Park (SENAPA), an interface where livestock and wildlife can interact, or current tick control practices in this area. During the dry season, pasture and water in the communal farmed areas of Serengeti District are in short supply and so cattle are grazed in the areas bordering the SENAPA (Laisser et al. 2014). The presence of buffalo within the national park, along with *Rhipicephalus appendiculatus*, can allow for cattle to become infected with *T. parva* while grazing (Marcellino et al. 2012). This system, therefore, provides an opportunity to assess the epidemiology of ECF in a livestock/wildlife interface area, by assessing the prevalence of *T. parva* in both cattle and buffalo and quantifying the current control measures being used.

This chapter describes a cross-sectional study with the following objectives:

2.2 Objectives

The overall aims of the project were to ascertain the prevalence and genetic and antigenic diversity of *T. parva* in both cattle and buffalo hosts, at the border of the Serengeti National Park. The specific objectives that are addressed in this chapter are to:

1. Estimate the prevalence of *T. parva* in the study area in cattle and buffalo and assess cow-level risk factors associated with the presence of *T. parva* infection
2. Identify control measures being used in the study area and investigate herd-level risk factors influencing uptake of control

2.3 Study area

The northern border of the Serengeti National Park (SENAPA), Northern Tanzania, was selected as it is an unfenced boundary area where livestock coexist with wildlife, including buffalo, providing opportunities for pathogens to cross over between populations (Figure 2.2). A significant sampling framework exists in and around the SENAPA, upon which the project has built.

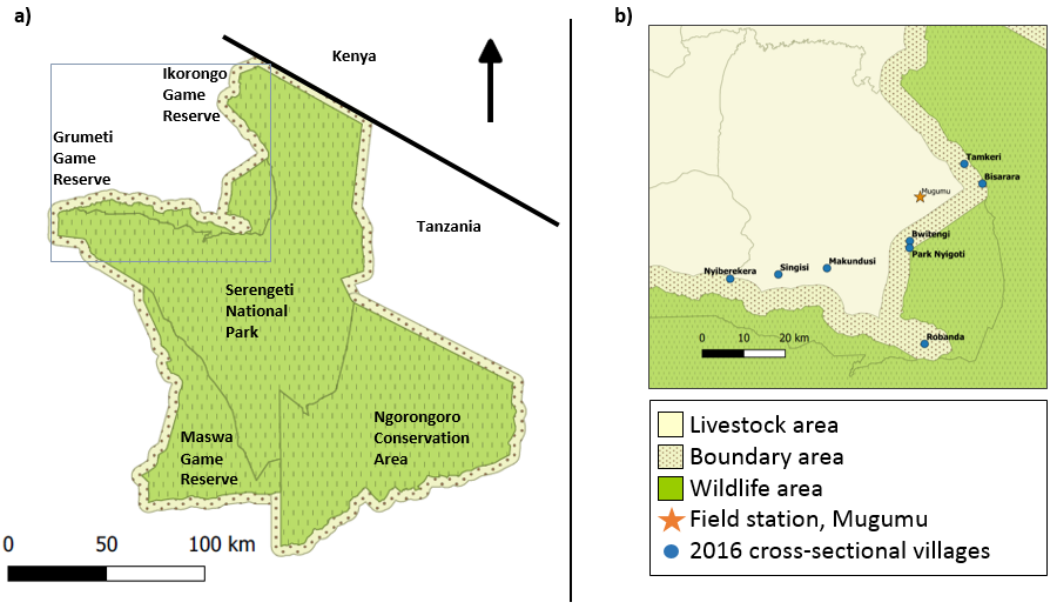

Figure 2.2: Map of study area
Showing a) Serengeti-Mara ecosystem with administrative areas marked; b) the study area showing wildlife (green), boundary (beige dotted) and livestock (cream) areas. The boundary area, highlighted as a 5 km buffer, is unfenced and is an interface where livestock can interact with wildlife. Villages where cattle sampling was carried out in 2016 shown in blue. The field station in Mugumu (indicated with an orange star) provided a base for the duration of the cattle sampling.

2.3.1 Location: Administrative areas

The Serengeti-Mara ecosystem extends over an area of 25,000 km^2 in Northern Tanzania, East Africa positioned at 34° - 36° Longitude and 1°30' - 3°30' Latitude. This ecosystem includes several administrative areas - Serengeti National Park (SENAPA) (14,763 km^2), Ngorongoro Conservation Area (8,288 km^2), as well as Maswa Game Reserve (2,200 km^2), Ikorongo and Grumeti Game Reserves and several Game Controlled Areas (Sinclair 1995) (Figure 2.2).

In 1929 an area of southern and eastern Serengeti was designated as a game reserve, followed by Protected Area status in 1940 and National Park status in 1951. The SENAPA was appointed status of World Heritage Site by UNESCO in 1981 (Sinclair 1995).

2.3.2 Climate

Serengeti has a tropical climate, with a mean monthly maximum temperature of 27°-28°C. The hotter months of October to March have a minimum temperature of 16°C and the cooler months of May to August a minimum of 13°C. Typical rainfall pattern is bimodal; long rains are from March until May and short rains from November to December. Rainfall pattern can, however, often be atypical with the rains merging into a single long duration, mostly in the north, or a failure of the short rains, especially in the southeast (Sinclair 1995).

2.3.3 Vegetation

The majority of vegetation in the Serengeti comprises savanna grasslands and acacia woodlands. A vegetation gradient mirrors the rainfall gradient, with longer grasses and woodland in the central, northern and western areas and shorter grass plains in the south east (Boone et al. 2006).

2.3.4 Ecosystem

The Serengeti ecosystem consists of a high density and diversity of wildlife, including buffalo. Grazing of livestock is not permitted in the protected National Park or the Game Reserves. The areas outside the protected areas are a mixture of agriculture and agropastoralism in the western border area, and in contrast is mostly pastoralist to the east. The study area is home to multiple tribes, including Sukuma, Kurya and Ikoma (Campbell and Hofer 1995). Due to the arid and semi-arid areas where pastoralists live, crop production is difficult in East Africa and so there is a lot of rural poverty and a real reliance on livestock for subsistence (Minjauw 2003).

The SENAPA is the conservation ‘core’ of the Tanzanian Serengeti-Mara ecosystem and the land in this area is solely for wildlife tourism (Homewood et al. 2001). This conservation core is surrounded by ‘buffer’ areas, consisting of the Ngorongoro Conservation Area (NCA), Loliondo Game Controlled Area

(LGCA), as well as Maswa, Grumeti and Ikorongo game reserves in Tanzania. Within the LGCA there is tourism, settlement, pastoralism, cultivation and licensed hunting. Land use in the NCA is for tourism, settlement, livestock herding and small-scale cultivation. In contrast the game reserves in Tanzania only allow land to be used for tourism and licensed hunting (Homewood et al. 2001).

2.3.5 Buffalo distribution

In the most recent aerial census a total of 55,411 buffalo were counted in the Serengeti ecosystem (TAWIRI 2014). Figure 2.3 shows buffalo distribution in the Serengeti ecosystem. Although the majority (97%) of buffalo were inside the protected areas, there were a few (3.0%) counted outside these areas (TAWIRI 2014) and it is important to note that the protected areas are not fenced but allow free movement of all wildlife, permitting interaction with livestock.

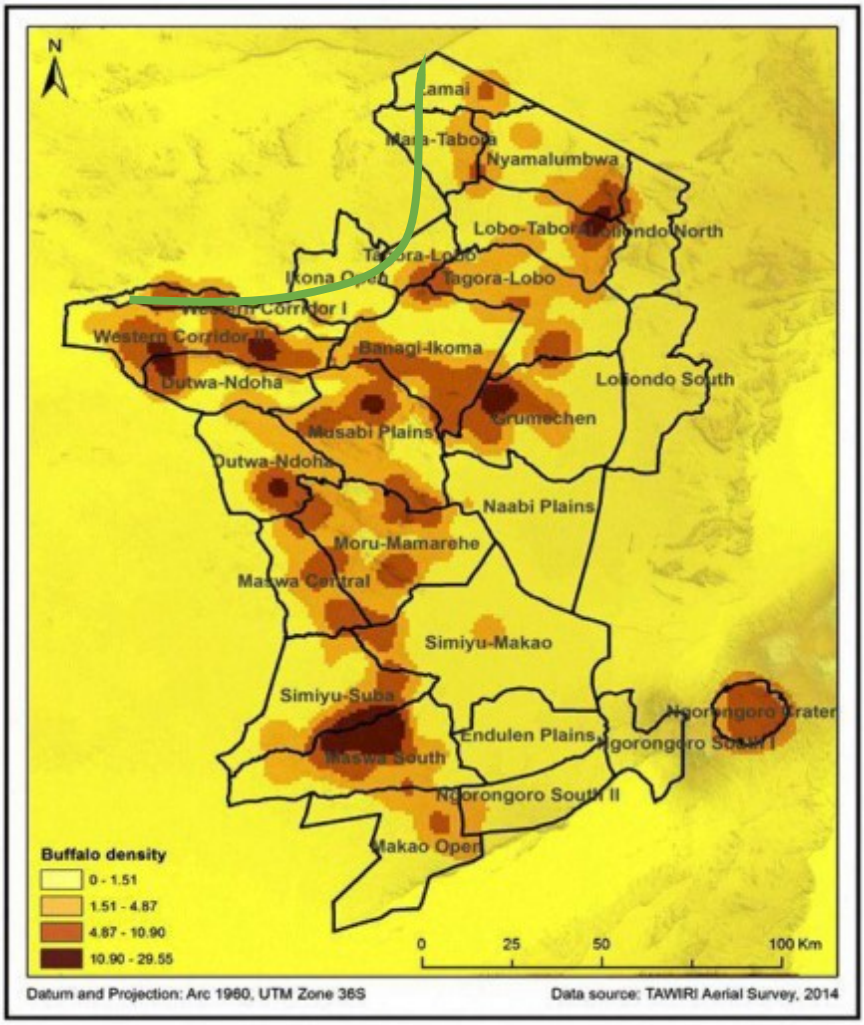

Figure 2.3: Map showing distribution of buffalo in the Serengeti ecosystem
Aerial census 2014 (buffalo density units not defined) (TAWIRI, 2014). The study area is along the northern border of the Serengeti ecosystem, shown in green.

2.3.6 Cattle distribution

Communities in the study area, northwest of the Serengeti National Park, practice livestock keeping as well as mixed crop-livestock farming (Estes et al. 2012). Native indigenous cattle breeds farmed in this area are predominantly Zebu x Tarime and Zebu x Maswa (Sahiwal, Boran and Mpwapwa cross breeds) (personal communication, Tito Kagize, Serengeti District Veterinary Officer). Livestock density is highest along the game reserve boundaries north-west and south-west of the National Park as shown in Figure 2.4.

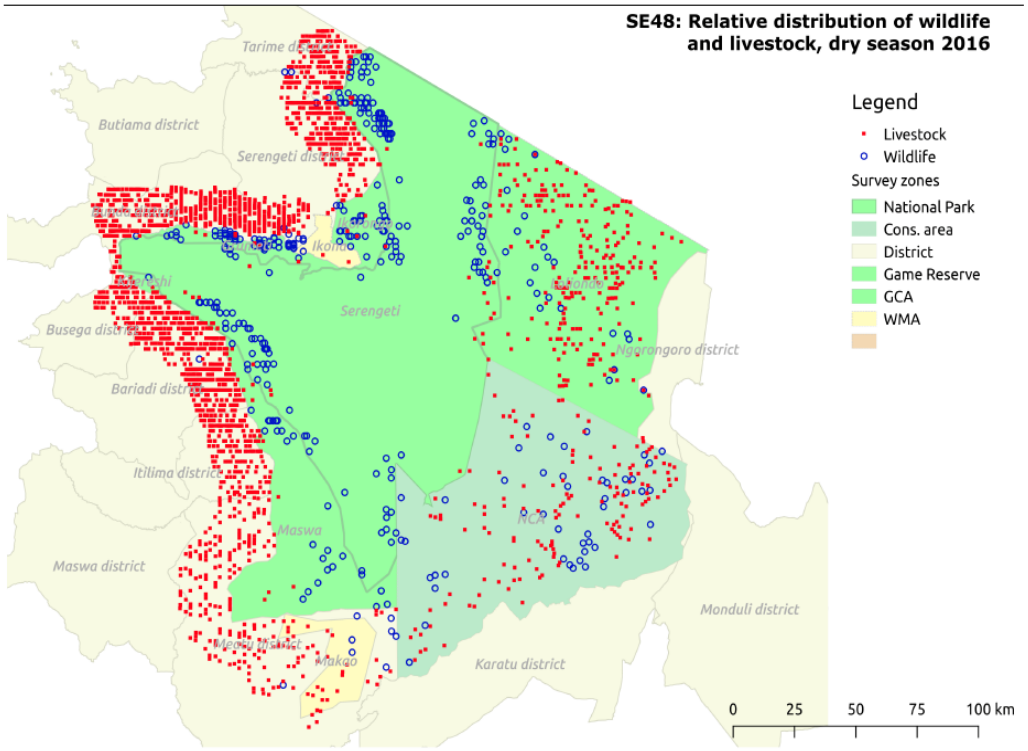

Figure 2.4: Map showing relative distribution of wildlife and livestock in the Serengeti ecosystem, dry season 2016 (TAWIRI 2016)

In the most recent aerial census carried out by TAWIRI (2016) a total of 1,210,846 ± 54,098 cattle were counted in the Serengeti ecosystem, with 40,251 ± 19,679 cattle estimated in the Serengeti National Park (3.3% of total counted), Ikorongo Game Reserve had an estimated 3,710 ± 2,006 (0.3%) and Grumeti Game Reserve had an estimate of 533 ± 431 (0.04%), despite the fact that cattle grazing is not permitted in protected areas. The survey area was

defined as a 10 km buffer inside and 20 km outside of SENAPA, as well as the entirety of Ngorongoro Conservation Area, Loliondo Game Controlled Area, and the whole of Maswa Game Reserve, plus a 20 km buffer outside of Maswa Game Reserve.

Serengeti District (Figure 2.5) specifically had 400,823 cattle in the 2017/2018 livestock registration campaign (personal communication, Tito Kagize, Serengeti District Veterinary Officer). The average herd size was 9 cattle (range 1-30) in the majority (87%) of cattle-rearing households (Ministry of Agriculture 2012).

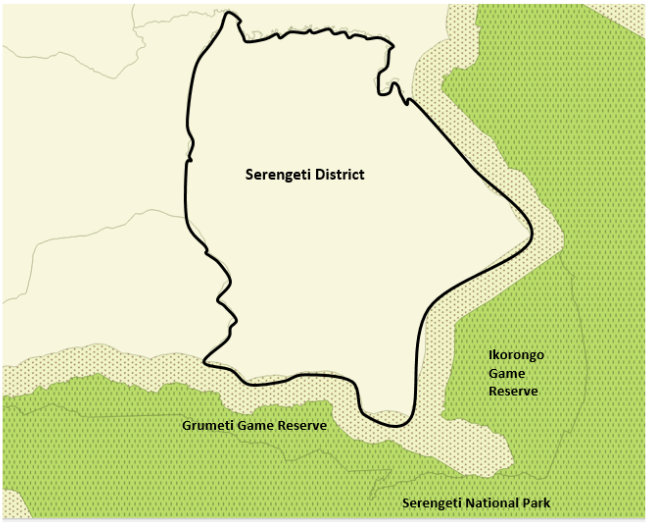

Figure 2.5: Map showing parts of Serengeti District outside protected areas Protected areas are also illustrated to demonstrate relationship between all areas

2.4 Materials and Methods

2.4.1 Introduction

To achieve the objectives identified in section 2.2 a series of surveys were designed and implemented:

1. Cross-sectional cattle sampling survey - to estimate prevalence of *T. parva* in cattle in the study area
2. Questionnaire survey of livestock owners - to establish vector-borne disease and vector control practices used by farmers
3. Buffalo sampling survey – to estimate prevalence of *T. parva* in buffalo at the study interface, using samples collected historically

2.4.2 Cattle Sampling Survey

A cross-sectional cattle sampling survey was designed and carried out in July and August 2016. The purpose of the survey was to gain an estimate of the prevalence of *T. parva* in cattle farmed at the boundary of the Serengeti National Park.

2.4.3 Ethical approval and permits – cattle and buffalo

The work in Tanzania was approved by Tanzania Wildlife Research Institute (TAWIRI) and Commission for Science and Technology (COSTECH) (Research Permit Number 2016-32-NA-2016-19) and ethical clearance for animal sampling as part of the cross-sectional survey was gained from Scotland's Rural College (SRUC) Animal Experimentation Committee.

2.4.4 Sampling strategy

The study design and sample size for the cross-sectional survey were aimed at establishing trypanosome prevalence in cattle (personal communication, Harriet Auty). Sample size was estimated for establishing the prevalence of *T. parva* to ensure the proposed survey would have sufficient power for this

purpose. A sample size calculation using an estimated prevalence of 20.0% (Swai et al. 2007), a precision of 5.0% and a 95% confidence interval, determined as:

$$n = \frac{1.96^2 \, P_{exp} \, (1 - P_{exp})}{d^2},$$

where n = required sample size; P_{exp} = expected prevalence; d = desired absolute precision, and 1.96 is the multiplier for a 95% confidence interval (Thrusfield 2005). Based on the parameters described, a sample size of 246 was calculated. The calculation does not allow for a clustering effect, where variation may be greater between clusters (herds) than within them. This reduces the effective sample size compared to a scenario with no clustering. To compensate for this the sample size was increased by a factor of two to allow for clustering, assuming a design effect of two, as is typical for vector-borne diseases in an African context (Donner and Klar 2000, Otte and Gumm 1997, Thrusfield 2005). This brought the required sample size to 502. The trypanosome study aimed to sample 700 cattle, so the sample size was sufficient to estimate *T. parva* prevalence at the desired precision. Point prevalence was determined as (Thrusfield 2005):

$$Prevalence\ (\%) = \frac{\text{Number of existing cases during specified time period x } 100}{\text{Population at risk during that specified time period}}$$

A multistage stratified sampling strategy was used to select herds (Wesonga et al. 2015). Regional areas in Tanzania are divided into districts. Districts are subdivided into local wards and, in rural areas, these are further subdivided into villages. Villages have subvillage divisions, for the purpose of administration over individual households. Eight villages within 5 km of the park boundary were selected; seven of the villages were randomly selected and one village, Robanda, was purposefully selected. Robanda was selected due to its location on the SENAPA boundary; it has a lot of wildlife area

surrounding it and has previously been found to have high prevalence of trypanosomiasis (Kaare et al. 2007). Within each of the eight villages, two subvillages were then selected at random. Subvillage authorities provided a sampling frame consisting of the list of livestock-owning households in the subvillage. Three herds were randomly selected from each subvillage by drawing numbers out of a hat. The stratified sampling strategy provided 48 selected herds for sampling, shown in Figure 2.6.

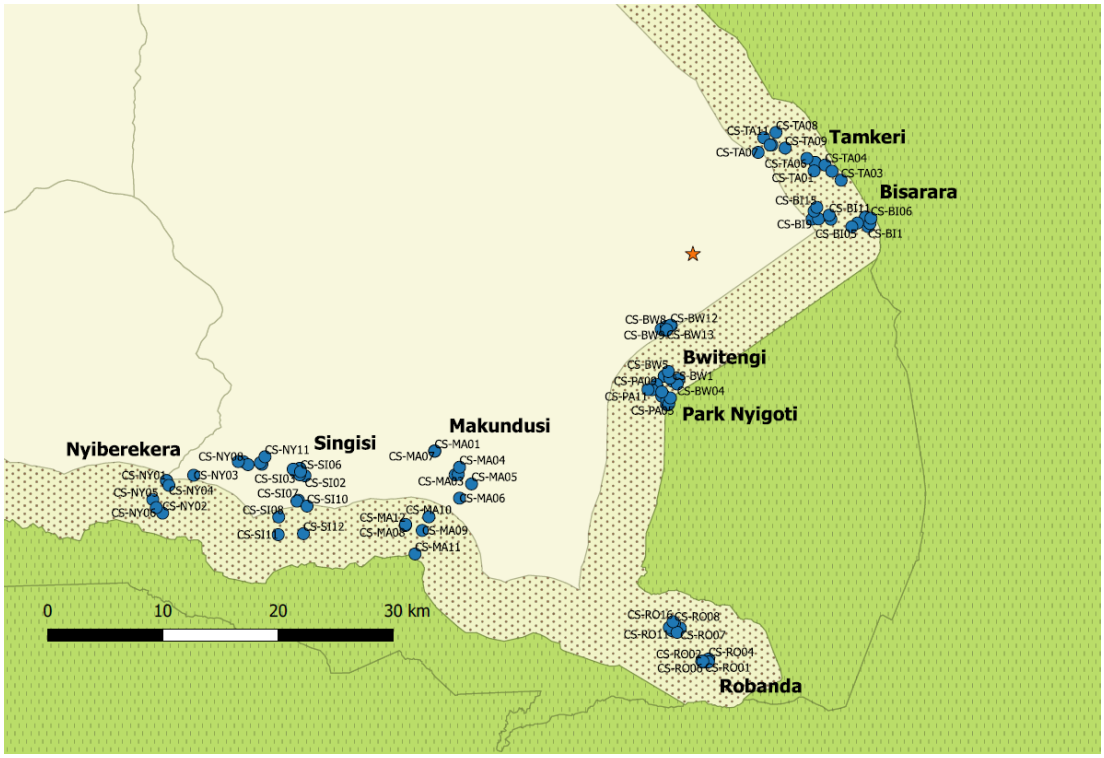

Figure 2.6: Map of cross-sectional herds sampled
The eight selected villages are labelled (bold text). Herds are shown in blue and labelled with herd identifying codes. Boundary of SENAPA shown in dotted beige. Wildlife area shown in green. Field station, Mugumu, shown as orange star.

In accordance with the sample size calculation a maximum of 20 cattle per herd were sampled, or all cattle if the herd size was fewer than 20. Cattle younger than 6 months were not included as farmers usually practice zero-grazing in order to protect them from disease (Muhanguzi et al. 2014a, Swai et al. 2009). The resulting final sample size was a total of 773 cattle sampled in the cross-sectional survey of 2016.

2.4.5 Sample collection

Sampling and handling of cattle was carried out by local Serengeti District livestock officers and a Tanzanian veterinarian. Cattle were gathered into the household boma enclosure and, as far as possible, were randomly selected for sampling. A 10ml blood sample was collected from the jugular vein into a PAXgene tube (PAXgene Blood DNA System, Qiagen). The owner of the sampled cattle was thanked with a sachet of trypanocide.

2.4.6 Covariate data collection

Cow-level data were collected for each animal sampled, including age, sex, origin (homebred or bought-in), health status at time of sampling (sick or healthy), record of any recent treatment and body condition score (BCS). Age was defined according to dentition (Kivaria et al. 2012). Body condition score was recorded numerically from 1-5, where 1=very poor and 5=very good. Informal discussion of these data indicated poor calibration, whereby some scores were recorded with decimal points, and so the decision was made to round the scores and group into more general categories of 'poor', 'fair' or 'good'. All covariate data were recorded by a single livestock officer to ensure consistency.

The location coordinates of each household were recorded by Global Positioning System (GPS) using a Garmin eTrex®10 handheld navigator.

2.4.7 Tick count

A half body tick count scoring system was created. To quantify ticks present on individual animals, each animal was examined by a livestock officer for ticks on the upper exposed half of the body, based on guidelines described by Walker et al. (2003). In order to minimise the period of time the animal was restrained, a tick count scoring system was created, similar to that described by Simuunza et al. (2011); a count was carried out for the half body as well as a separate count for the ears, the predilection site for *Rhipicephalus*

appendiculatus. The categories established for tick counts were category 0 (0 ticks), category 1 (1-10 ticks), category 2 (11-50 ticks) and category 3 (>50 ticks).

2.4.8 Questionnaire Survey - Purpose

A structured questionnaire survey was designed to gather specific information on the herds recruited in the cross-sectional study in July and August 2016, in order to assess vector control practices being used in the SENAPA study site, to look at herd level risk factors for vector control, as well as informing the analysis of *T. parva* prevalence in samples from these farms. The questionnaire was part of a wider survey that included the collection of data on animal African trypanosomiasis (AAT) and tsetse fly vectors; these data were to be used in parallel projects and have not been analysed as part of this thesis. The questionnaire was also administered to herds in an ongoing longitudinal study (n=23) that will be discussed in section 2.4.15.

2.4.9 Questionnaire Survey - Topics

The key topics covered in the questionnaire were:

- Farmer demographic information and herd management (global positioning system (GPS) location of farm, number of cattle owned, where cattle kept, other species kept, cattle movements)
- Farmer knowledge of vectors (ability to identify ticks and tsetse, awareness of vectors on/around cattle and farm, seasons when vectors seen)
- Vector control methods (methods of application, products being used, seasons products used, dilution, volume and dose of products used, frequency of use and reasons for use)
- Knowledge of East Coast fever (clinical signs, cause, cases, prevention and treatment).

2.4.10 Questionnaire Ethical review

A proposal for the questionnaire was submitted to the University of Edinburgh Human Ethical Review Committee, along with a participant consent form and participant information sheet. The questionnaire and all accompanying documents were approved for distribution. Certification of ethical approval is included in Appendix A.

2.4.11 Questionnaire Translation

As the questionnaire was to be completed by Tanzanian farmers, it was written in English but translated into the local language Kiswahili by a native Kiswahili-speaker and experienced translator. It was then back-translated into English, also by a native Kiswahili-speaker to ensure it was conveying the correct meaning. As the participants had variable levels of literacy, the questionnaire was designed to be read out in an interview-style and completed by a native Kiswahili speaking enumerator, experienced in conducting similar questionnaires. In the administration of a questionnaire there is the potential for bias as the interviews could influence judgement if a question was further explained or elaborated (Choi and Pak 2005), but the risk was minimised by the use of an experienced enumerator, by using the same enumerator for all questionnaires, and by instigating a training period to ensure the enumerator understood the importance of consistency in asking questions.

2.4.12 Questionnaire Participant consent

A participant consent form and information hand-out/supplement were given to each participant. Both these documents were provided in Kiswahili and were also explained verbally to every participant. Consent to partake in the questionnaire survey was given by the participant in the format of signature or thumbprint. Participants were provided with an oxytetracycline spray (for livestock) to thank them for their contribution and time.

2.4.13 Questionnaire Pilot study

The questionnaire was piloted with two farmers known to the enumerator (and outside the cross-sectional and longitudinal sampling herds described above). Minor amendments were made. Ideally a larger pilot study would have been conducted, but the pilot size was constrained by logistical issues of time availability in the study site.

2.4.14 Questionnaire design

In designing the questionnaire it was important that the respondent would be able to understand the question, able to answer the question, and willing to provide the information needed to answer the question. Therefore, it was necessary to keep each question limited so as not to appear burdensome (Jepson et al. 2005). The questions were to be clear and concise so as to be easily understood and the questions were written with the intention of being unbiased and therefore not influencing the respondents answers (Dohoo 2009). The questionnaire comprised 64 questions. Seven questions were of a closed multiple-choice format, all allowing other answers to be given if not listed. Two questions required the respondent to identify an image. Eighteen questions were binary, requiring “yes” or “no” responses. The remaining questions were open-ended requiring free text answers. The questionnaire is included in Appendix B.

2.4.15 Questionnaire distribution

Sample size calculations for questionnaires are difficult given the range of question types asked. Sample size calculation was carried out for a yes/no question (e.g. do you spray your cattle?) based on an estimated proportion of 50% (this requires the highest sample size, so gives the most conservative estimate), desired precision of 10% and 95% confidence intervals, which gives a sample size of 97.

Cross-sectional herds (n=97) were randomly selected in 2016, 48 of which also had their cattle sampled, as already described. As part of this sampling strategy, an additional 49 herds were randomly selected from the same villages and subvillages as those from which the sampled cattle inhabited. Cattle in these additional herds were not sampled but the farmers were interviewed with the questionnaire survey.

In addition, 23 herds already involved in a longitudinal study were asked to complete the questionnaire. These herds were recruited in 2013 (prior to this project), initially in a study looking at Foot and Mouth Disease virus (FMDV) and subsequently in projects analysing AAT. These herds were repeatedly sampled approximately every six to nine months. The samples came from identifiable cattle (individually ear-tagged) so provided a longitudinal sample set of approximately 4000 samples and the sampling is ongoing.

For the cross-sectional herds, an average of 6 interviews were carried out daily over a three week period. Each questionnaire interview lasted on average one hour. For the longitudinal herds, an average of three interviews were carried out daily over a 10 day period, taking into consideration longer distances involved in visiting these farms and the generally poor conditions of the roads. On completion of the questionnaires in Swahili, they were carefully translated into English. Herds were assigned an identifying code in order to allow farmer identity to be anonymised. The raw data was then entered into Microsoft Excel and was then uploaded into the open source statistical analysis software R (R Core Team 2018).

2.4.16 Calculations

In order to calculate the distance between each farm and the boundary of the protected area, the household GPS locations were used in combination with geographical shapefiles in a script (Mazeri 2019) in open source statistical analysis software R (R Core Team 2018).

To calculate the dose of acaricide being used, data were collected on the preparation of acaricides. Dose was based on the dilution that the farmers were making up the products and the volume of diluted product that they were applying to the herd, divided by cattle number per herd. Using this information then allowed the dose-per-cow to be calculated and compared to the prescribed guidelines found on most (but not all) packaging; for Albadip there was no recommended volume stated on the packaging. The manufacturer was contacted several times but never responded and so the volume advised to farmers by the local livestock officer was used. Standardised "correct" dose was taken to be ≥ 1 (with under-dosing < 1).

2.4.17 Buffalo Sampling Survey

Buffalo samples were collected from the Serengeti National Park (n=22) in 2011 (Figure 2.7). A power calculation was carried out on the sample size to establish the level of precision; this was based on an estimated prevalence of 90% (Oura et al. 2011a, Pienaar et al. 2011, Young et al. 1978), 95% confidence intervals and sample size of 22, and provided a level of 10% precision, sufficient for the purpose of this survey. These samples were kindly collected opportunistically by veterinary colleagues working in the SENAPA when buffalo were darted for other purposes.

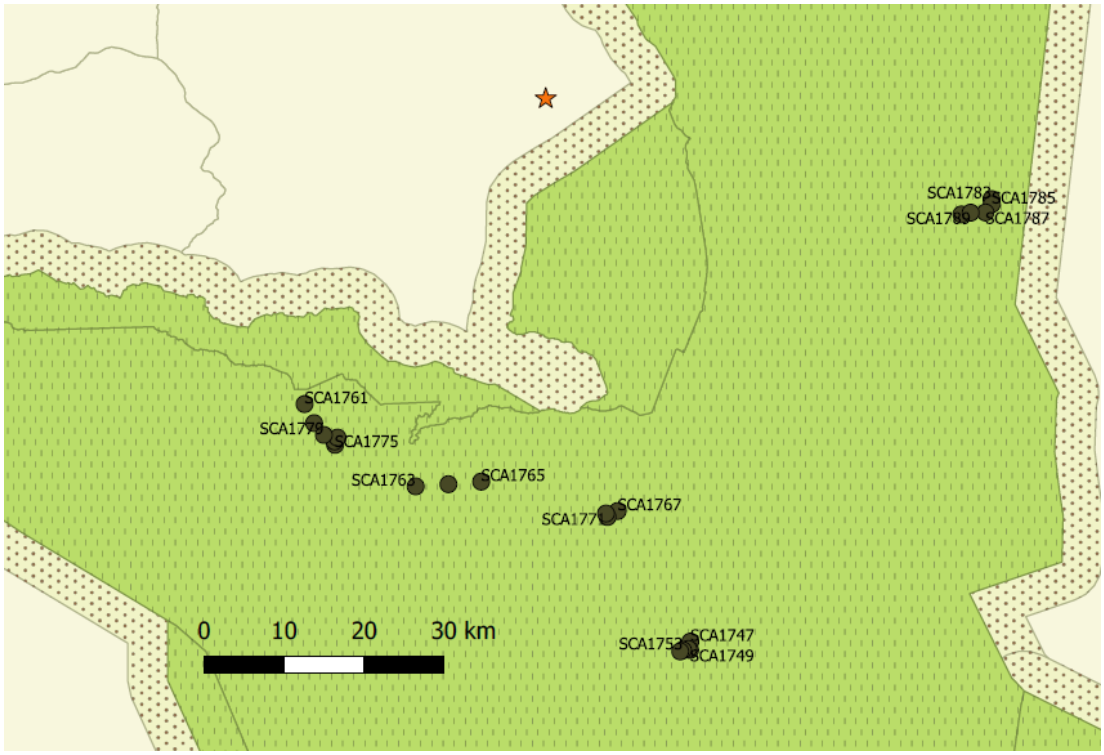

Figure 2.7: Map of sampled buffalo in Serengeti National Park
Buffalo shown in brown. Boundary shown in dotted beige. Wildlife area shown in green. Field station in Mugumu shown as orange star.

2.4.18 PCR detection of *T. parva*

Sample preparation and processing, including DNA extraction and PCR assay methodology, is described in detail in Chapter 3. Briefly, DNA was subjected to a nested PCR, targeting *T. parva* p104 (Odongo et al. 2010, Skilton et al. 2002), using primers IL3231 and IL755 to amplify a 496 bp fragment followed by primers p104_f2_Od and p104_r2_Od to amplify a 277 bp fragment. Cycling conditions for round one were 94°C for 1 minute, 40 cycles (94°C for 1 minute, 60°C for 1 minute, 72°C for 1 minute) followed by 72°C for 9 minutes. Round two conditions were 94°C for 1 minute, 30 cycles (94°C for 1 minute, 55°C for 1 minute, 72°C for 1 minute) followed by 72°C for 9 minutes. A sample was considered positive when a band was visible upon gel electrophoresis of PCR amplicons.

2.4.19 Statistical analysis of sampling and questionnaire surveys

PCR prevalence data, covariate data and questionnaire datasets were initially entered into Microsoft Excel (Microsoft Corporation, USA) before being imported into R (R Core Team 2018). The 'tidyverse' (Wickham 2017) suite of packages was used to clean data; this involved programmatic coding of missing data and categorising continuous variables. Datasets were joined programmatically before being analysed using packages and functions in R. Variables were explored and visualised by creating histograms and scatterplots in order to get an initial impression of the data as well as detecting erroneous data. Boxplots were created using R function 'boxplot' to display the distribution of continuous variables, showing the first (lower) and third (upper) quartiles, the median and the lower and upper whiskers and outliers (Chambers et al. 1983, Murrell 2005). Package 'ggplot2' was used to produce graphics. QGIS® version 2.18.28 was used to create maps (unless otherwise cited) using shape files and GPS coordinates. Prevalence of *T. parva* was established for the 2016 cross-sectional survey, at village-level and herd-level. A random effects model was used to estimate the overall and village level prevalences. This model adjusts the estimates and confidence intervals for the clustering of animals within villages and subvillages which reduces the effective sample size and hence increases the width of the confidence intervals. The mixed effects models were run using the 'LME4' R package.

A univariable risk factor analysis was carried out to check for any confounding factors by fitting generalised linear models, using Wald test. Logistic regression models were used to establish cow-level risk factors (predictors) of being PCR positive for *T. parva* as well as analysis of herd-level risk factors for vector control. Cow-level variables considered for the model were age, sex, origin, body condition score, half-body tick count, ear ticks and treatment in the past six months, and analysis was performed at the univariable level only. This statistical model uses a logistic function to model a binary dependent variable. R function 'glm' was used. Odds ratios (OR) with 95% confidence interval (CI) and p-value were calculated for all variables. Binomial confidence intervals were calculated using Clopper-Pearson method (Clopper

and Pearson 1934, Conover 1971) and using R function 'binom.test'. Where generalised linear models could not be fit, Firth's regression and Fisher's Exact test were used. Firth's regression is used when there is low or high prevalence, or "separation", which can be problematic for fitting models (Fijorek 2012, Rahman and Sultana 2017) and R function 'logistf' was used. Fisher's Exact test, using R function 'fisher.test', compares the proportions for nominal variables in a two by two table and is appropriate for small sample sizes (Agresti 2002). Likelihood ratio test (LRT) comparisons were used to assess the significance of overall variables of being PCR positive for *T. parva*, by assessing the goodness of fit of two competing models, one without the variable and the other model with the variable included (Buse 1982), using R function 'lrtest'. Numerical outcomes were modelled using univariable linear regression. Statistical tests were considered significant at $p \leq 0.05$.

2.5 Results

2.5.1 Overall prevalence of *T. parva* in cattle and buffalo

A total of 773 cattle from eight villages, 16 subvillages and 48 herds were sampled in the 2016 cross-sectional survey. Three samples had to be removed from the sample set due to mixed labelling of samples and missing covariate data. Of the 770 cattle, 39 were positive by PCR for *T. parva.* Overall raw prevalence of *T. parva* in the cattle survey was 5.07% (CI: 3.70-7.00%). Prevalence adjusted for clustering was 5.20% (CI: 2.80-7.70%).

Buffalo samples (n=22) were all infected with *T. parva* (100.00%, CI: 85.00-100.00%).

2.5.2 Prevalence of *T. parva* in cattle at village level

Prevalence of *T. parva* at village level (Table 2:2) was variable, ranging from 0.00% (Robanda) to 8.77% (Singisi) (Figure 2.8). Prevalence adjusted for clustering was calculated for all villages except Robanda, and ranged from 2.50% (Bisarara) to 9.60% (Singisi). Robanda was significant with zero

prevalence (p = 0.019) but there was no significant difference in prevalence of *T. parva* between the other villages.

Table 2:2: Prevalence of *T. parva* distibution by village

Village	Cattle number	Raw prevalence	Adjusted prevalence and 95% CI
Bwitengi	113	7/113 (6.19%)	6.60% (3.10-14.20)
Bisarara	81	2/81 (2.47%)	2.50% (0.60-10.30)
Makundusi	115	6/115 (5.22%)	5.50% (2.40-12.50)
Nyiberekera	91	6/91 (6.59%)	8.10% (3.50-18.60)
Park Nyigoti	92	3/92 (3.26%)	3.40% (1.10-10.90)
Singisi	114	10/114 (8.77%)	9.60% (5.00-18.40)
Tamkeri	77	5/77 (6.49%)	7.20% (2.90-18.00)
Robanda[+]	88	0/88 (0.00%)	NA

[+]Due to zero prevalence and therefore complete separation of the data in Robanda, it was not possible to include Robanda in the generalised linear model and so Fisher's Exact test was used for this village instead. It was also not possible to include Robanda in the Random effects model for adjusted prevalence

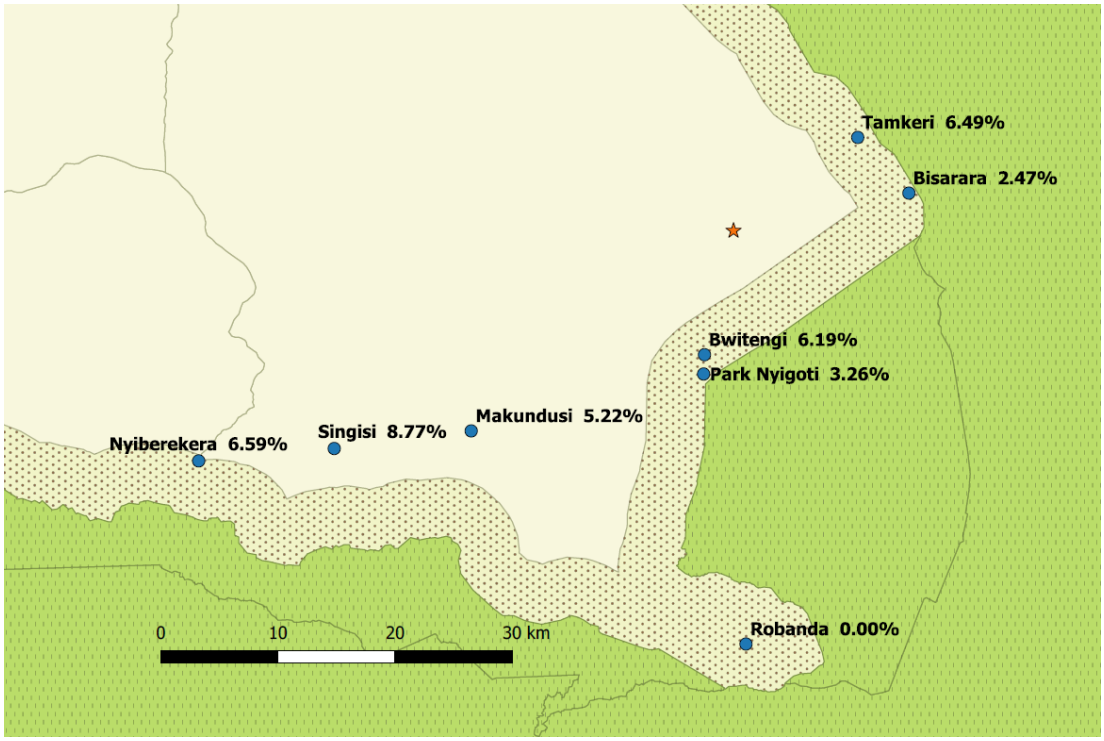

Figure 2.8: Map showing *T. parva* prevalence in sampled villages
Villages shown in blue. Field station, Mugumu shown as orange star.

2.5.3 Prevalence of *T. parva* in cattle at herd level

The prevalence of *T. parva* varied from 0.00% to 16.67% at herd level (Figure 2.9). Twenty two of the forty eight herds had zero infection detected. These herds were from a variety of villages but notably all herds from Robanda village had zero prevalence. Herd NY07, from Nyiberekera village, and TA07, from Tamkeri village, had the highest prevalence of 16.67%. Herd was not found to be a significant risk factor for *T. parva* prevalence (LRT 52.4, p = 0.274). None of the differences in individual herd prevalence were found to be significant.

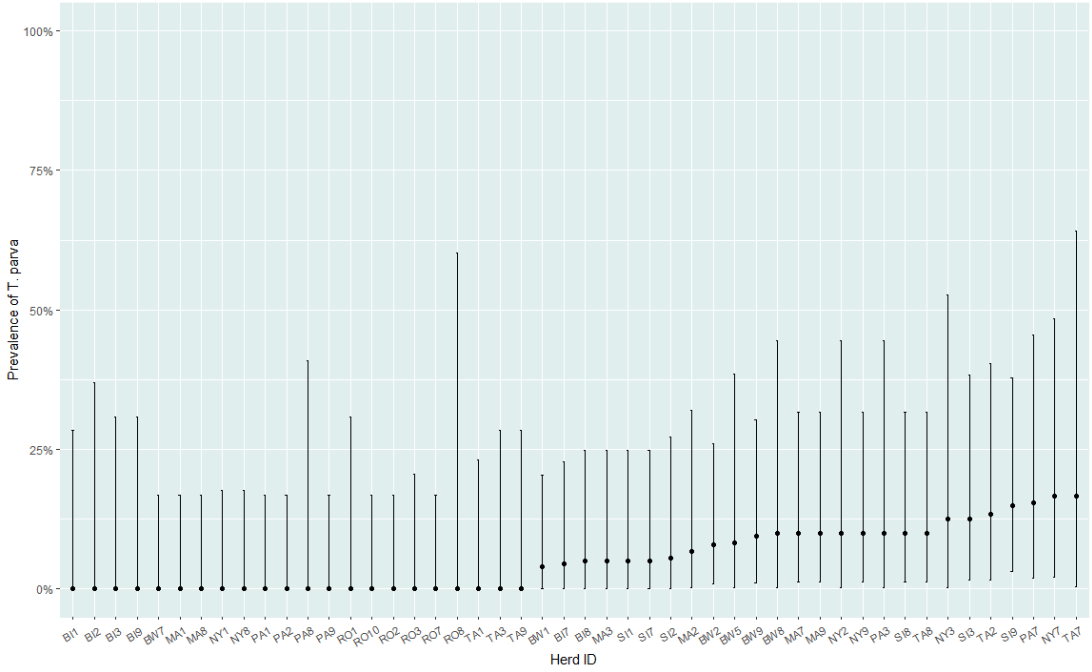

Figure 2.9: Point prevalence of *T. parva* at the herd level
including (95%) binomial exact confidence interval.

2.5.4 Questionnaire descriptive analysis

As well as providing the covariate data to inform interpretation of the results of molecular analyses and samples from the cross-sectional cattle herds, the questionnaire results reveal more general information about ECF and vector control in the study area. In the cross-sectional survey, herds were randomly selected; these herds are therefore representative of the population at the interface. In contrast, due to the self-selection component of the longitudinal

farmer recruitment (farmers volunteered to participate), the farmers of these herds have a tendency to be proactively concerned with health interventions, also being more knowledgeable and therefore biased towards better education and wealth than the 'average' farmer.

Analysis was carried out on all questionnaires, cross-sectional and longitudinal combined (n=120). Due to their differences as stated, the cross-sectional and longitudinal questionnaires were also analysed separately. Results will be presented for both datasets combined and separately. The response rate per question is stated throughout the results.

2.5.5 Profile of participants

A total of 120 people were recruited, as described in 2.4.15, to participate in the questionnaire. Of those, 97 farmed cross-sectional herds (80.83%) and 23 participants (19.16%) farmed longitudinal herds.

Most participants were male (77.5%, 93/120) and 65% of participants (78/120) were the head of the household.

2.5.6 Cattle demographics

All participants provided information on the number of cattle they owned, representing a total of 9,122 animals. In the longitudinal herds cattle numbers ranged from 46 to 1000 (mean 218.6, median 150) and in the cross-sectional herd numbers ranged from 4 to 280 cattle per herd (mean 42.2, median 25) (Figure 2.10). A Wilcox test was conducted in R (wilcox.test) to compare herd-size in the different surveys, cross-sectional and longitudinal. There was a significant difference ($p = <0.0001$) in the mean herd sizes.

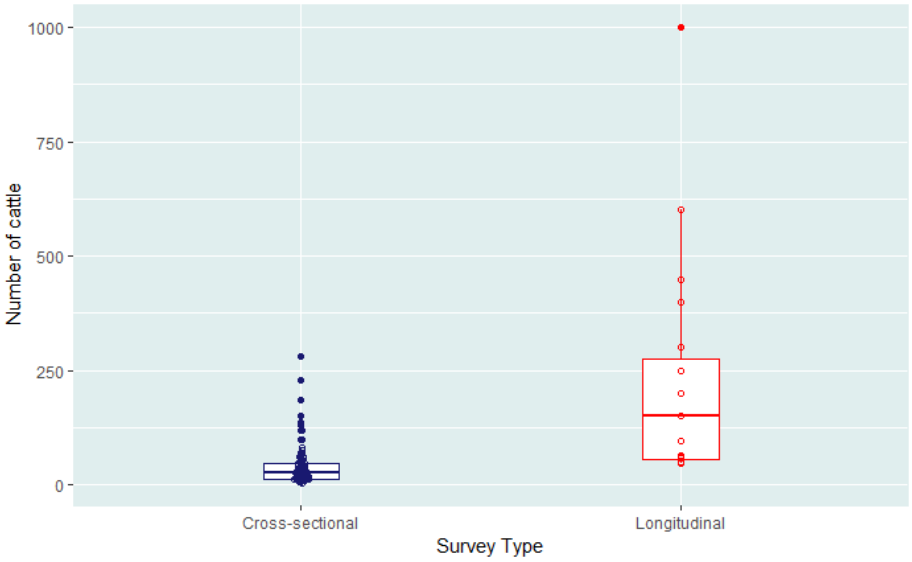

Figure 2.10: Cattle herd sizes
Boxplot of cross-sectional (n=97) cattle herds (blue) and longitudinal (n=23) cattle herds (red), showing the lower quartile, the median and the upper quartile as horizontal lines and vertical lines indicating variability outside the lower and upper quartiles. Outliers are shown as individual points.

Most of the farmers (58.33%, 70/120) owned all of the cattle on their farm, with 41.6% (50/120) having cattle also belonging to someone else (58.76% cross-sectional, 56.52% longitudinal).

When asking farmers about the origin of their cattle, 82.5% (99/120) reported their cattle were homebred (78.35% cross-sectional, 100% longitudinal).

2.5.7 Cattle movement

All farmers reported keeping their cattle at their farm overnight i.e. returning nightly from grazing or watering destinations. As agro-pastoralists, the farmers in the questionnaire survey routinely move their cattle for grazing and watering. Farmers reported travelling between zero and eight kilometres daily for grazing in the wet season (mean 1.35, median 1) (0.05-6km, mean 1.23, median 1 cross-sectional; 0-8km, mean 1.90, median 1.5 longitudinal) and 0-22km daily in the dry season (mean 3.43, median 2.5) (0.05-10km, mean 2.74, median 2 cross-sectional; 0-22km, mean 6.35, median 4 longitudinal). Between zero

and seven kilometres were travelled daily for water in the wet season (mean 1.38, median 1) (0.1-3km, mean 1.25, median 1 cross-sectional; 0-7km, mean 1.94, median 2 longitudinal) and 0.3-22km were travelled daily for water in the dry season (mean 3.47, median 2.5) (0.3-7km, mean 2.81, median 2.5 cross-sectional; 0.6-22km, mean 6.27, median 4 longitudinal).

Approximately one quarter of farmers (24.17%, 29/120) reported sending their cattle away for periods of time, for grazing or watering, as draught or milking animals and for weaning (10.31%, 10/97 cross-sectional; 82.61%, 19/23 longitudinal). The length of time that cattle were sent away ranged from zero to twelve months (mean 2.5, median 2) (1-2 months, mean 3.2, median 2 cross-sectional, 0-5 months, mean 2.05, median 2 longitudinal).

The months that farmers sent their cattle away varied from June through to December, but predominantly were August, September and October (August and September cross-sectional; September and October longitudinal) (Table 2:3).

Table 2:3: Months farmers send their cattle away

Month sent away	June	July	August	September	October	November	December
All participants	15.78% (3/19)	31.57% (6/19)	63.15% (12/19)	89.47% (17/19)	63.15% (12/19)	15.78% (3/19)	10.52% (2/19)
Cross-sectional participants	11.11% (1/9)	33.33% (3/9)	88.88% (8/9)	88.88% (8/9)	44.44% (4/9)	11.11% (1/9)	11.11% (1/9)
Longitudinal participants	20% (2/10)	30% (3/10)	40% (4/10)	90% (9/10)	80% (8/10)	20% (2/10)	10% (1/10)

When asked how far cattle travelled when sent away, farmers reported sending them between 0-45km (mean 9.63, median 7) (0-20km, mean 10.18, median 12 cross-sectional; 0-45km, mean 8.22, median 7 longitudinal).

Farmers were asked about the presence of sheep or goats; 77.5% farmers kept sheep on their farms (93/120) (79.38%, 77/97 cross-sectional and 69.56%, 16/23 longitudinal) and 74.2% farmers kept goats (89/120) (71.13%,

69/97 cross-sectional and 86.96%, 20/23 longitudinal). Sheep numbers (Figure 2.11a) ranged from 0-300 per farm (mean 32.49, median 11) (0-300, mean 27.75, median 10 cross-sectional; 0-300, mean 52.48, median 20 longitudinal). Goat numbers (Figure 2.11b) ranged from 0-180 per farm (mean 19.38, median 10) (0-180, mean 15.85, median 6 cross-sectional; 0-100, mean 34.6, median 28 longitudinal).

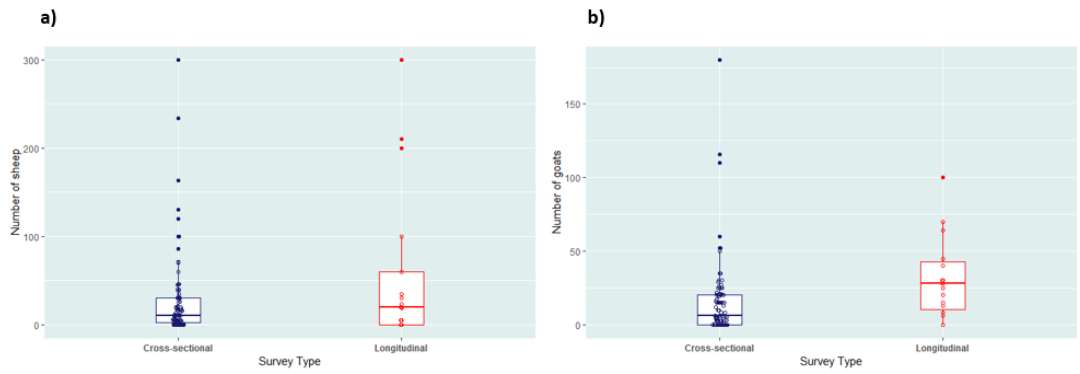

Figure 2.11: Sheep and goat flock sizes

a) Sheep flock sizes for cross-sectional (n=97) (blue) and longitudinal (n=23) (red) farms and
b) b) Goat flock sizes for cross-sectional (n=97) (blue) and longitudinal (n=23) (red) farms.

2.5.8 Knowledge of vectors

Farmers were shown a series of five different insect vector images and asked to identify a tick. Of the 120 farmers interviewed 118 were able to correctly identify a tick (98.33%) (97.93%, 95/97 cross-sectional and 100% longitudinal). All farmers reported seeing ticks on their cattle. Farmers were asked about which body parts they saw ticks on their cattle. The body parts most reported were udder (89.16%), breast (74.16%), scrotum (72.50%), ears (56.66%) and axilla (50.83%) (Table 2:4).

Table 2:4: Cattle body parts on which farmers reported seeing ticks

Body part	Udder	Breast	Scrotum	Ears	Axilla	Belly	Anus	Tail
All participants	89.16% (107/120)	74.16 (89/120)	72.50% (87/120)	56.66% (68/120)	50.83% (61/120)	25.0% (30/120)	17.50% (21/120)	11.66% (14/120)
Cross-sectional participants	88.65% (86/97)	73.19% (71/97)	83.50% (81/97)	53.6% (52/97)	53.6% (52/97)	27.85% (27/97)	17.52% (17/97)	10.30% (10/97)
Longitudinal participants	91.30% (21/23)	78.26% (18/23)	26.08% (6/23)	69.56% (16/23)	39.13% (9/23)	13.04% (3/23)	17.39% (4/23)	17.39% (4/23)

Farmers reported seeing ticks on their cattle all year round, with peak times being June until September (Figure 2.12), the dry season.

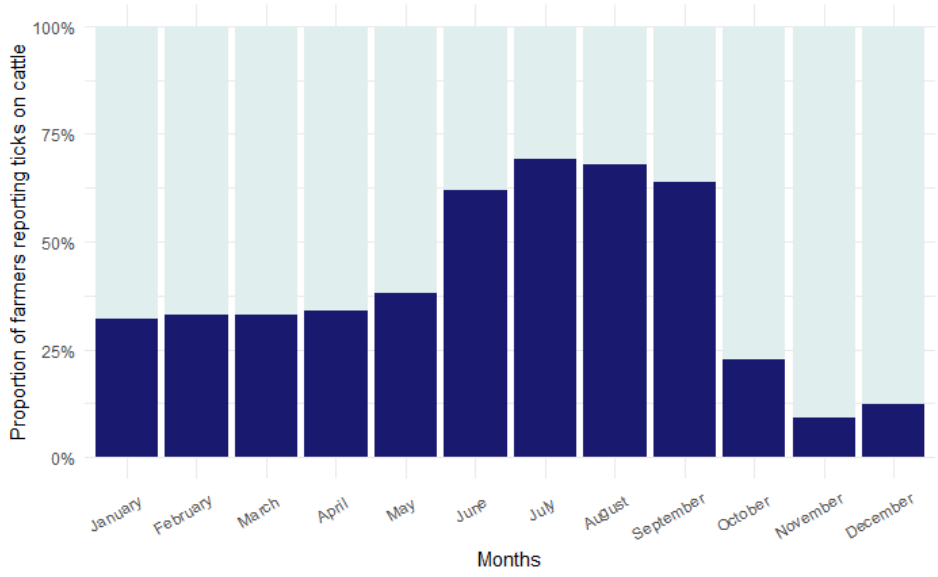

Figure 2.12: Months when farmers reported seeing ticks on their cattle.
Peak months are June-September.

When asked about locations that farmers saw ticks on their cattle they mostly reported seeing them while at grazing and did not see them on their farms.

As well as being asked broadly about which body parts farmers saw ticks, they were asked in a separate question if they saw ticks on the ears of cattle specifically. This gave a different response, with 83.76% farmers reporting seeing ticks on the ears of cattle (98/117) (82.29%, 79/96 cross-sectional;

90.47%, 19/21 longitudinal), and these were mostly seen around grazing areas and not on farms.

Farmers were asked if they knew what diseases ticks transmitted; only 23.07% of all farmers thought that ticks transmitted East Coast fever (27/117) (14.89% 14/94 cross-sectional; 56.52%, 13/23 longitudinal). Farmers also reported other diseases (or clinical signs) that they associated with ticks (Table 2:5).

Table 2:5: Conditions farmers reported being transmissted by ticks

Condition	high fever	ECF	anaplasma	heartwater	babesia	enlarged lymph nodes	anaemia	do not know
All participants	62.93% (73/116)	23.07% (27/117)	13.79% (16/117)	12.93% (15/115)	4.31% (5/115)	1.72% (2/116)	0.862% (1/116)	16.23% (19/116)
Cross-sectional participants	73.11% (68/93)	14.89% (14/94)	10.75% (10/93)	9.67% (9/93)	1.07% (1/93)	1.07% (1/93)	1.07% (1/93)	15.95% (15/94)
Longitudinal participants	21.73% (5/23)	56.52% (13/23)	26.08% (6/23)	26.08% (6/23)	17.39% (4/23)	4.34% (1/23)	0.00 (0/23)	17.39% (4/23)

Approximately two thirds (62.93%) of farmers (73.11%, 68/93 cross-sectional and 21.73%, 5/23 longitudinal) recognised 'high fever' as a clinical sign that they associated with ticks. Longitudinal farmers generally reported more awareness of the other common tick-borne diseases (TBDs), namely, anaplasmosis, ehrlichiosis (heartwater) and babesiosis.

2.5.9 Vector prevention

Almost all farmers (99.16%, 119/120) reported doing some form of tick prevention (98.96%, 96/97 cross-sectional; 100%, 23/23 longitudinal). Of the farmers, 97.5% (117/120) were using prevention products (98.96%, 96/97 cross-sectional; 91.30%, 21/23 longitudinal). Two farmers (1.66%) (0%, 0/97 cross-sectional; 8.69%, 2/23 longitudinal) reported hand-picking as tick prevention.

The majority of farmers (86.66%) (104/120) reported doing tick prevention all year round (83.50%, 81/97 cross-sectional, 100%; 23/23 longitudinal), with

12.5% (15/120) only doing so when “ticks were bad” (15.46%, 15/97 cross-sectional; 0%, 0/23 longitudinal).

When asked about how they were applying tick prevention products, 79.16% (95/120) farmers were spraying their cattle (74.22%, 72/97 cross-sectional; 100%, 23/23 longitudinal) and 40.83% (49/120) farmers were dipping their cattle (42.26%, 41/97 cross-sectional; 34.78%, 8/23 longitudinal).

Farmers were asked how they applied spray products; 75.83% (91/120) farmers described spraying “all over the body” (74.22%, 72/97 cross-sectional; 82.60%, 19/23 longitudinal). Some of the longitudinal farmers described specific body areas that they sprayed; 13.04% (3/23) udder, 13.04% (3/23) axilla, 13.04% (3/23) ears, 8.69% (2/23) back and 8.69% (2/23) belly.

The time interval for spraying cattle ranged from every 4 to 270 days (mean 21.2, median 10) (4-270 days, mean 24.44, median 10 cross-sectional; 4-30 days, mean 10.65, median 7.5 longitudinal) (Figure 2.13). Six farmers reported spraying cattle ‘when they saw ticks’ or ‘when ticks were bad’ (4/97 cross-sectional and 2/23 longitudinal) rather than at a predetermined interval.

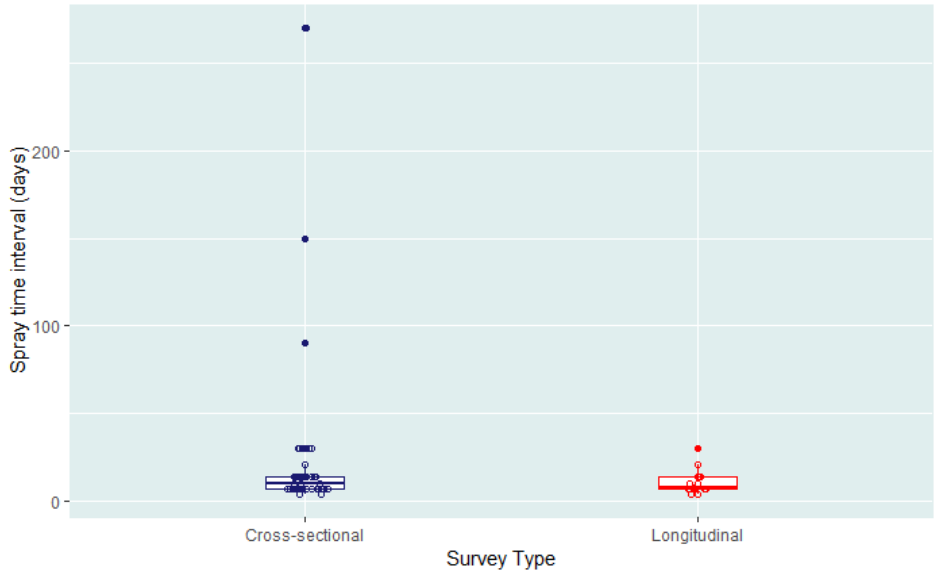

Figure 2.13: Spray time intervals
Cross-sectional (blue) and longitudinal (red) farmers.

For the purpose of analysis, spray time intervals were put into similar sized classes. The most common time interval class that farmers sprayed their cattle was every 5-7 days (37.89, 36/95) (38.89%, 28/72 cross-sectional; 34.78%, 8/23 longitudinal) followed by every 8-14 days (28.42%, 27/95) (26.38%, 19/72 cross-sectional; 34.78%, 8/23 longitudinal) (Figure 2.14). A smaller proportion of farmers sprayed every 29-35 days (12.63%, 12/95) (15.2%, 11/72 cross-sectional; 4.35%, 1/23 longitudinal) and 4.21% sprayed (4/95) sprayed every four days (2.78%, 2/72 cross-sectional; 8.70%, 2/23 longitudinal). A few cross-sectional farmers sprayed at intervals greater than every 50 days (5.55%, 4/72).

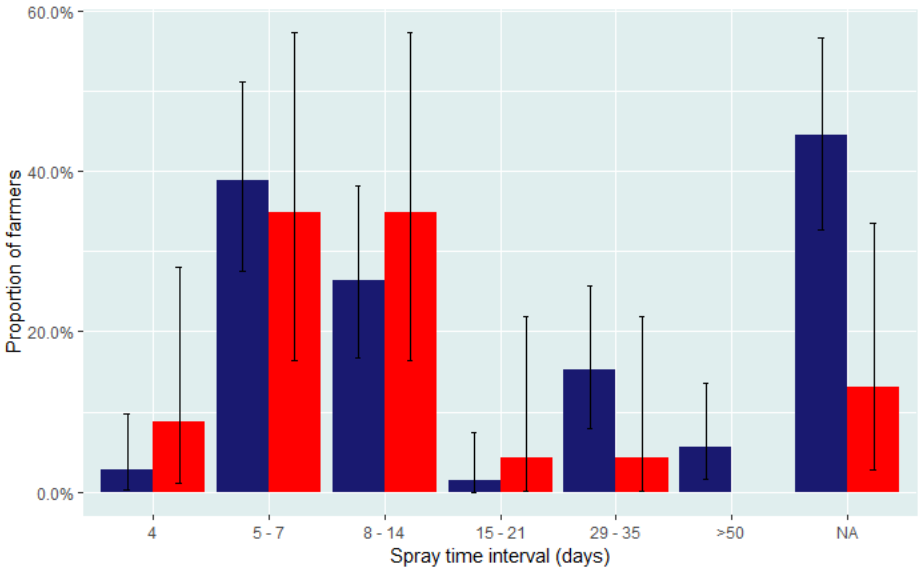

Figure 2.14: Time interval of acaricide spraying
Cross-sectional (blue) farmers and longitudinal (red) farmers. 'NA' indicates non-responses.

The time interval for dipping cattle ranged from every 7 to 150 days (mean 24.40, median 14 days) (7-150 days, mean 25.33, median 14 cross-sectional and 14-60 days, mean 19.75, median 14 longitudinal) (Figure 2.15).

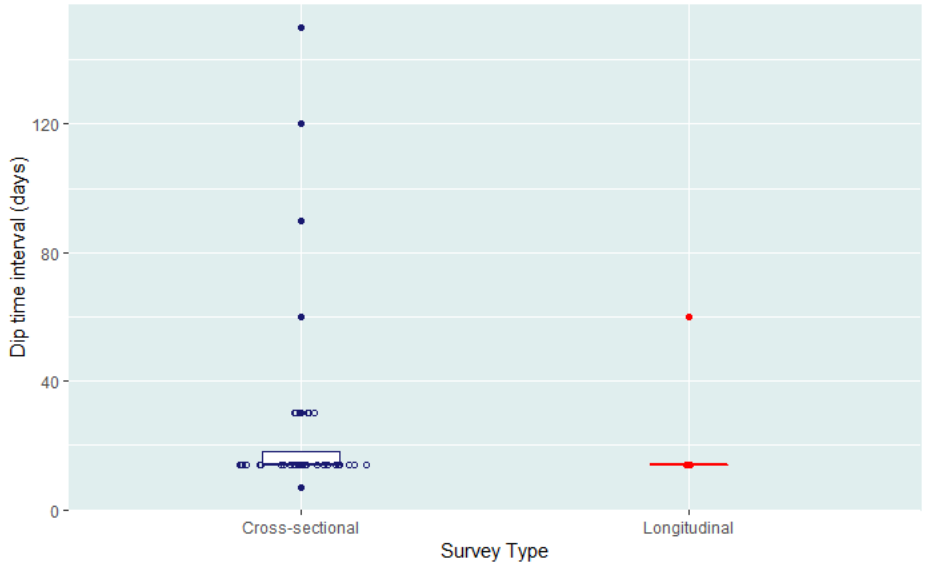

Figure 2.15: Dip time intervals
Cross-sectional (blue) and longitudinal (red) farmers.

Of the farmers that were dipping their cattle, the majority reported doing so every 8-14 days (73.5%, 36/49) (70.73%, 29/41 cross-sectional and 87.50%, 7/8 longitudinal) followed by every 29-35 days (12.24%, 6/49) (6/41, 14.63% cross-sectional and 0.00%, 0/8 longitudinal) and greater than every 50 days (10.20%, 5/49) (9.76%, 4/41 cross-sectional; 12.5%, 1/8 longitudinal) (Figure 2.16).

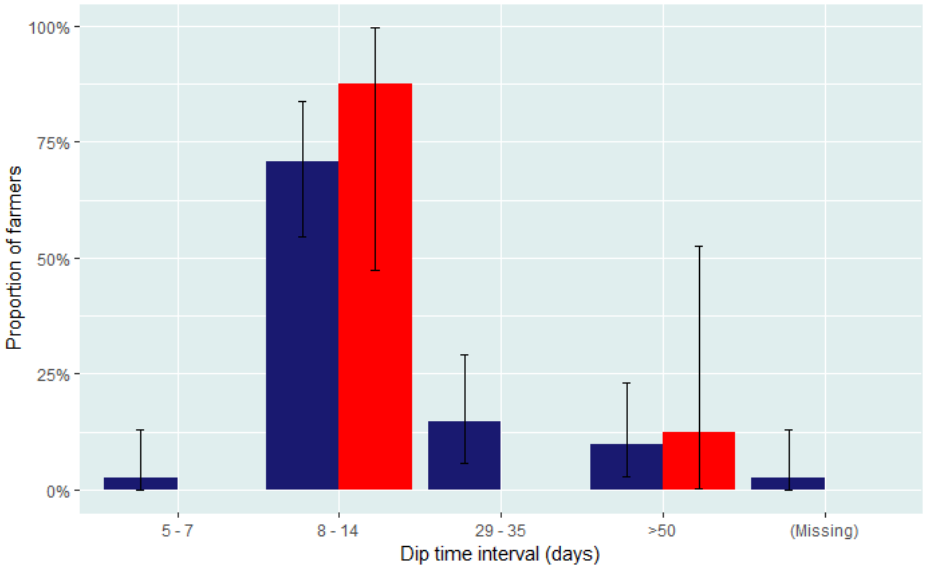

Figure 2.16: Time interval of acaricide dipping
Cross-sectional (blue) farmers and longitudinal (red) farmers.

2.5.10 Knowledge of ECF

Almost half the farmers (45.83%, 55/120) had heard of East Coast fever (35.05% (34/97) cross-sectional, 91.30% (21/23) Longitudinal). One third (34.74%, 41/118) thought they knew the clinical signs associated with ECF (23.15% (22/95) cross-sectional, 82.60% (19/23) Longitudinal). A range of clinical signs considered attributable to ECF were reported (Figure 2.17).

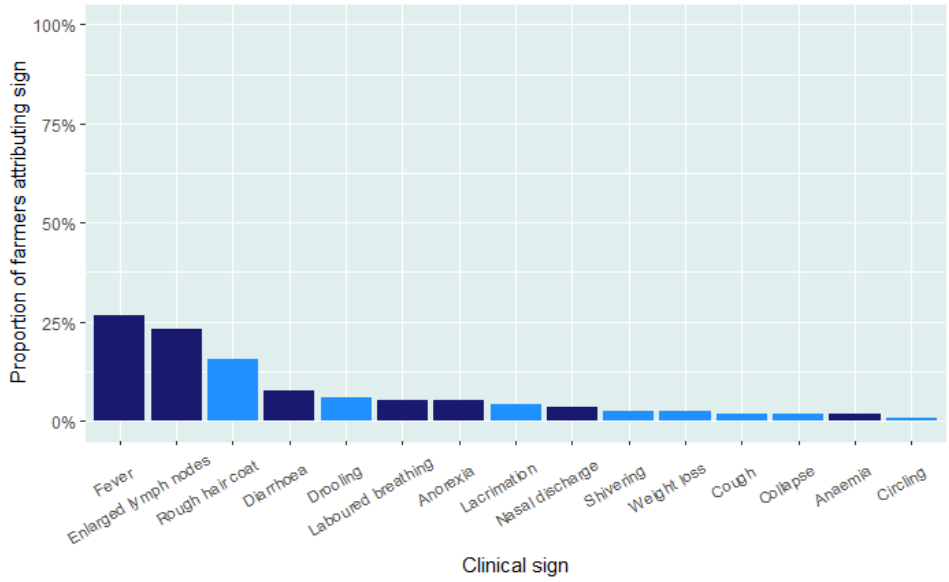

Figure 2.17: Clinical signs farmers attribute to ECF
Clinical signs that are correct are highlighted in dark blue. Signs reported by farmers that are not typically associated with ECF are shown in lighter blue.

When asked if farmers knew that ticks caused ECF 29.16% (35/120) of farmers knew that ticks caused ECF (20.61%, 20/97 cross-sectional; 65.21%, 15/23 longitudinal) and 22.12% (25/113) farmers reported having cases of ECF in their cattle (5.55%, 5/90 cross-sectional; 86.95%, 20/23 longitudinal). Farmers were asked about how many cases of ECF they had in the past one year. They reported between 0 and 60 cases (mean 6.25, median 1.5) (1-12 cases, mean 1.66, median 0 cross-sectional; 0-60 cases, mean 10.29, median 4 longitudinal). Of the cases of ECF that farmers reported to have had, the range of numbers of deaths apparently attributable to ECF was 0 to 12 (mean 1.59, median 0) (0-9 deaths, mean 1.2, median 0 cross-sectional; 0-12 deaths, mean 1.94, median 1 longitudinal).

Farmers were asked if they used prevention methods for ECF. Of the 38 farmers who responded, 32 (84.21%) reported using prevention methods (83.33%, 15/18 cross-sectional; 85% 17/20 longitudinal). Of those who responded, most stated they sprayed their cattle to prevent ECF, followed by a combination of both spraying and dipping (Table 2:6). A small number of cross-sectional farmers used oxytetracycline as a prevention method for ECF.

Table 2:6: ECF prevention methods reported by farmers

	Spraying	Spraying & Dipping	Dipping	Oxytetracycline
All participants	64.70% (22/34)	14.7% (5/34)	11.76% (4/34)	2.94% (1/34)
Cross-sectional participants	40% (6/15)	26.66% (4/15)	26.66% (4/15)	6.66% (1/15)
Longitudinal participants	84.21% (16/19)	5.26% (1/19)	0% (0/19)	0% (0/19)

No farmer reported using the ECF vaccine (Muguga Cocktail ITM) – 40 of the 120 participants reported not using it (20/97 cross-sectional; 20/23 longitudinal), with the remaining 80 participants not answering. When asked why they were not using the vaccine, the farmers stated that they did not know there was a vaccine (37/40, 18 cross-sectional, 19 longitudinal), they did not know about ECF (1 cross-sectional), or the vaccine was too expensive (1 longitudinal). Farmers reported varying treatments that they used for suspected cases of ECF (Table 2:7).

Table 2:7: ECF treatments reported by farmers

	OTC[1] 10%	OTC 20%	OTC 30%	OTC unspec.[2]	Hilet[3] 12%	Butalex[4]	Veterinary advice
All participants	59.37% (19/32)	65.62% (21/32)	40.62% (13/32)	15.62% (5/32)	6.45% (2/31)	9.67% (3/31)	6.45% (2/31)
Cross-sectional participants	75% (9/12)	100% (12/12)	83.33% 10/12)	8.33% (1/12)	0.00% (0/12)	0.00% (0/12)	0.00% (0/12)
Longitudinal participants	50.00% (10/20)	45.00% (9/20)	15.00% (3/20)	20.00% (4/20)	10.52% (2/19)	15.78% (3/19)	10.52% (2/19)

[1]OTC = oxytetracycline, [2]unspec. = unspecified percentage, [3]Hilet = oxytetracycline, [4]Butalex = buparvaquone

Of the 32 farmers who indicated they were using prevention methods, all reported using oxytetracycline, of varying concentration, to treat suspected cases of ECF. A few longitudinal farmers (15.78%) used Butalex (buparvaquone) and 10.52% sought veterinary advice.

Farmers purchased the ECF treatments predominantly from local private agroveterinary shops.

2.5.11 Products being used for tick prevention

Farmers described four tick prevention products that they sprayed their cattle with, namely Albadip, Paranex, Cybadip and Tantix, with some farmers reporting use of more than one (Table 2:8). All four products are synthetic pyrethroids (cypermethrins).

Table 2:8: Proportion of farmers using acaricide products

Product	Chemical	Cross-sectional farmers using	Longitudinal farmers using
Albadip	Alphacypermethrin 10%	43/79 (54.43%)	14/23 (60.86%)
Paranex	Alphacypermethrin	32/79 (40.51%)	12/23 (52.17%)
Cybadip	Cypermethrin 15%	11/79 (13.92%)	6/23 (26.08%)
Tantix	High-cis Cypermethrin 10%	5/79 (6.33%)	0/23 (0.0%)

Farmers were asked for details of how they used the acaricide products, specifically how they diluted the product and how much diluted product they applied to each animal. Dose per cow data is presented for Albadip (Figure 2.18), Paranex (Figure 2.19), Cybadip (Figure 2.20) and Tantix (Figure 2.21).

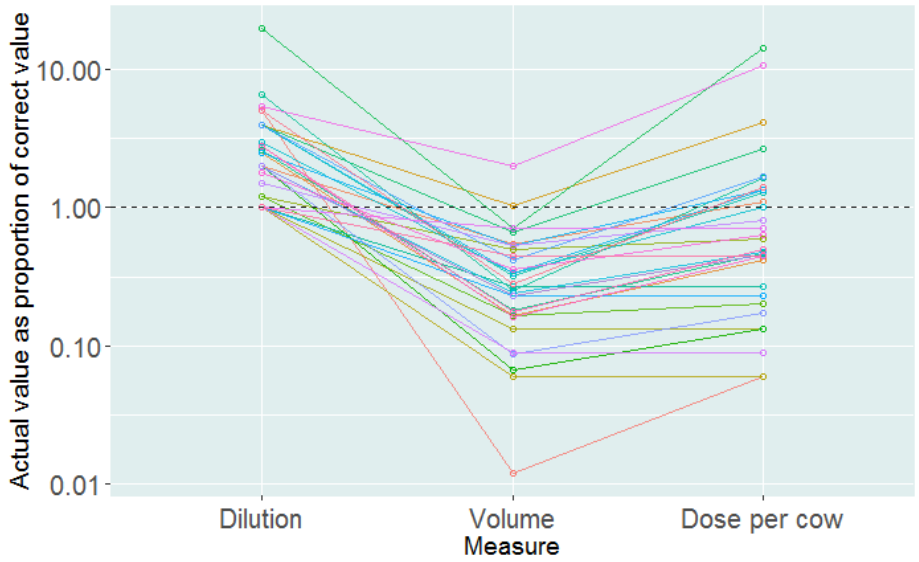

Figure 2.18: Scatterplot showing dilution, volume used and dose per cow of Albadip acaricide
Each herd shown as a proportion (log scale) of correct amount, where values < 1.0 represent underdosing (grey dashed line). Different coloured points represent different herds and lines connect individual herds.

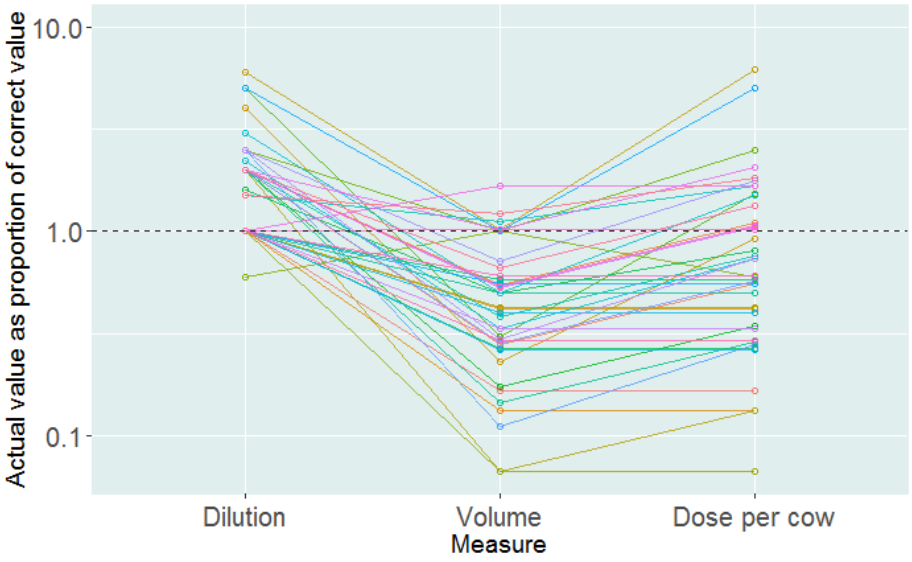

Figure 2.19: Scatterplot showing dilution, volume used and dose per cow of Paranex acaricide,
Each herd shown as a proportion (log scale) of correct amount, where values < 1.0 represent underdosing (grey dashed line). Different coloured points represent different herds and lines connect individual herds.

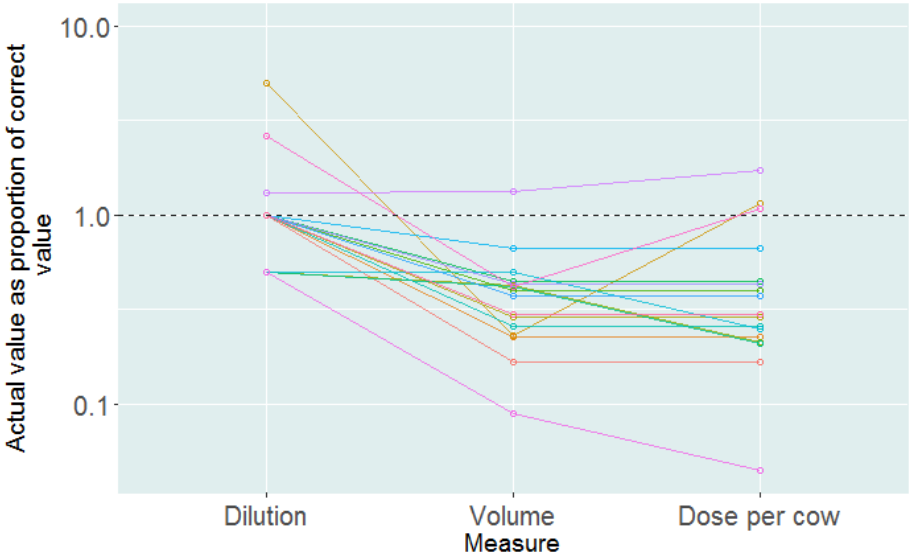

Figure 2.20: Scatterplot showing dilution, volume used and dose per cow of Cybadip acaricide,
Each herd shown as a proportion (log scale) of correct amount, where values < 1.0 represent underdosing (grey dashed line). Different coloured points represent different herds and lines connect individual herds.

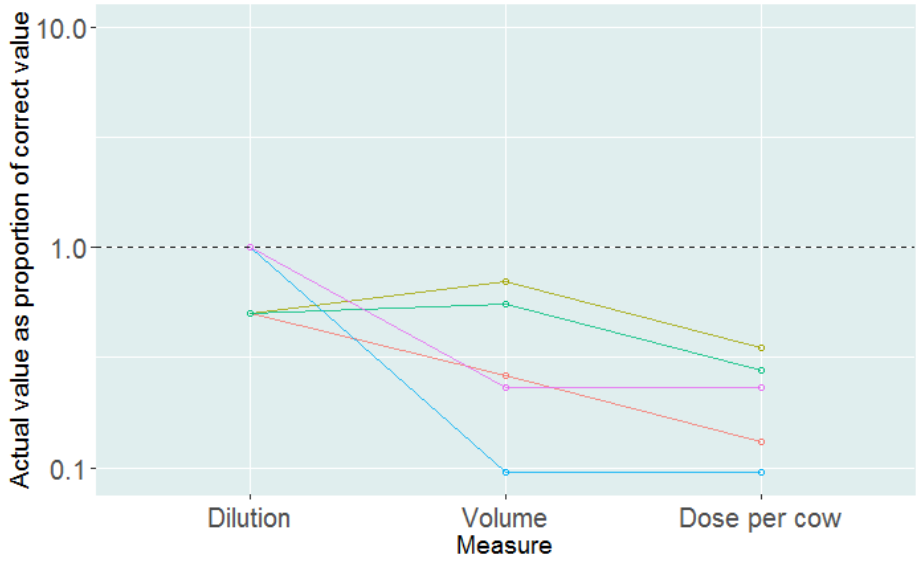

Figure 2.21: Scatterplot showing dilution, volume used and dose per cow of Tantix acaricide,
Each herd shown as a proportion (log scale) of correct amount, where values < 1.0 represent underdosing (grey dashed line). Different coloured points represent different herds and lines connect individual herds.

A low proportion of farmers were found to be dosing correctly, ranging from 0.00% dosing correctly with Tantix, 18.18% correctly using Cybadip, followed by 41.94% using Paranex correctly and Albadip showing best compliance with 43.48% dosing correctly (Table 2:9).

Table 2:9: Proportion of farmers dosing cattle with acaricide correctly

Product	Number farmers using product	Number farmers with dose data	Proportion correct dosing
Albadip	43	23	10/23 (43.48%)
Paranex	32	31	13/31 (41.94%)
Cybadip	11	11	2/11 (18.18%)
Tantix	5	5	0/5 (0%)

Farmers predominantly reported purchasing their acaricide products from the local agroveterinary shop, with two farmers reporting they purchased acaricide at the livestock markets. Albadip reported cost was between 20,000-35,000 Tanzanian Shillings (TZS) (mean 26,000, median 25,000), Paranex reported

cost was 35,000-70,000 TZS (mean 47,651.16, median 50,000), Cybadip reported cost was 20,000-30,000 TZS (mean 24,062.5, median 22,500) and Tantix reported cost was 24,000-40,000 TZS (mean 31,800, median 30,000) for one litre (where 3000 TZS is equivalent to approximately one British Pound (GBP)). To put this in context, the purchase cost of a cow in Serengeti district is between 200,000 and 600,000 TZS and a calf 100,000 to 250,000 TZS (personal communication, Livestock Officer Emmanuel Sindoya).

The same four cypermethrin acaricides were used for dipping. As the dip tanks are community-run, the acaricide products are made up by Livestock Officers and so the farmers did not know details on how the products were used. Farmers paid an arranged price of 100 TZS (personal communication, Emmanuel Sindoya, Serengeti District Livestock Officer) per cow for the use of the dip and so did not know costs of dip products or where they were purchased.

Farmers were asked if the acaricides they were using had any other benefits; 61.86% (73/118) (58.94%, 56/95 cross-sectional; 73.91%, 17/23 longitudinal) believed they repelled flies, 67.79% (80/118) thought they repelled tsetse (65.26%, 62/95 cross-sectional; 78.26%, 18/23 longitudinal) and 4.16% (5/120) did not know of other benefits (3.09%, 3/97 cross-sectional; 8.69%, 2/23 longitudinal).

2.5.12 Descriptive analysis of covariate data

In the cross-sectional survey, 561 cattle were female and 208 were male (the sex of one animal was not recorded). Cattle ages ranged from 6 months to eleven years old (mean 3.00, median 3.00). For the purpose of analysis, cattle ages were put into similar sized classes. The majority of cattle were reported to be homebred (82.86%).

Body condition score was coded as either ‘poor’, ‘fair’ or ‘good’. The majority of cattle were in ‘poor’ body condition (44.03%) with 39.74% in ‘fair’ body

condition and only the remaining 15.45% considered to be in 'good' body condition.

All cattle were reported healthy at the time of sampling. Some cattle (40.39%) had received a form of treatment in the past six months. Treatments reported were: multivitamin injection, Berenil® (diminazene aceturate), Novidium (homidium chloride), Ivermectin, Samorin (isometamidium chloride) Albendazole, Penicillin-Streptomycin and oxytetracycline.

2.5.13 Cow-level risk factor analysis

The association between individual cow-level factors and the odds of *T. parva* infection are shown at the univariable level (Table 2:10); potential risk factors included cattle age, sex, origin (homebred or bought-in), body condition score, half-body tick count, ear tick count and treatment in the past six months.

Table 2:10: Cow-level factors associated with *T. parva* prevalence based on logistic regression

Variable	Factor level	Total	*T. parva* positive (%)	OR	95% CI	LRT value	p-value
Age						5.5	0.019*
Age category (years)	0.5-1.5	191	7/189 (3.70%)	REF			
	1.5–2.5	164	4/164 (2.44%)	0.65	0.016-2.19		0.498
	2.5-3.0	113	9/113 (7.96%)	2.25	0.81-6.46		0.118
	3.0-4.5	175	5/175 (2.86%)	0.76	0.22-2.44		0.652
	4.5-11	119	14/119 (11.76%)	3.47	1.39-9.39		0.009 *
Cattle sex						0.35	0.554
	Male	208	9/208 (4.32%)	REF			
	Female	561	30/559 (5.37%)	1.25	0.61-2.85		0.561
Origin						0.62	0.429
	Bought-in	129	6/129 (4.65%)	REF			
	Homebred	636	33/636 (5.19%)	1.45	0.61-4.29		0.449
Body condition score						1.63	0.202
	Poor	339	16/339 (4.72%)	REF			
	Fair	304	18/304 (5.92%)	1.27	0.63-2.56		0.498

Variable	Factor level	Total	*T. parva* positive (%)	OR	95% CI	LRT value	p-value
	Good	119	5/119 (4.20%)	0.89	0.28-2.31		0.816
Half body tick count						0.15	0.699
	No ticks (0)	600	32/599 (5.34%)	REF			
	Few ticks (1-10)	160	6/159 (3.77%)	0.69	0.26-1.58		0.423
	Some ticks (11- >50)	9	1/9 (11.11%)	2.22	0.12-12.61		0.460
Ear ticks*						2.98	0.084
	Present	28	0/28 (0.00%)	REF			
	Absent	738	39/738 (5.28%)	25.1	0.04-67132		1.00
Treatment in past 6 months						0.002	0.961
	No treatment	456	23/431 (5.06%)	REF			
	Treatment	311	16/311 (5.15%)	1.02	0.52-1.94		0.961

*Due to complete separation of the data, Firth's regression was used to calculate the OR for Ear ticks and *T. parva* prevalence

Cattle age was found to be significant, with the older cattle (4.5-11 years) having increased risk (OR = 3.47, p = 0.009) of being positive by PCR for *T. parva* compared to the youngest cattle (reference level).

The number of ticks seen on the sampled cattle was very low with 77.92% cattle having a tick count of zero. Tick counts on cattle ears were correspondingly very low with only 3.66% cattle having ticks observed at this predilection site for *R. appendiculatus.* None of the cattle with ticks on their ears were positive for *T. parva*.

Cattle age was found to be the only significant risk factor for *T. parva* prevalence in the univariable analysis.

2.5.14 Herd-level risk factor analysis

Although not the aim of the study, herd level factors were also assessed for association with *T. parva* prevalence. Univariable analysis was carried out and results are presented in Appendix C. None of these twenty six variables investigated were found to be significantly associated with *T. parva* prevalence.

Acaricide dose was also investigated as a risk factor for prevalence; for Cybadip and Tantix it was not possible to do logistic regression as there were no cows with prevalence data that were receiving the correct dose (>1). The Odds Ratio for Paranex correct dose and prevalence was 1.49 (p = 0.894) and for Albadip correct dose and prevalence was 1.13 (p = 0.970), i.e. no significant association between correct dose and prevalence of *T. parva*.

2.5.15 Herd-level risk factors influencing uptake of vector control

Herd-size was investigated as a possible influence for vector control options and it was also investigated for association with awareness of ECF (Table 2:11). All questionnaires were combined for this analysis, cross-sectional and longitudinal.

Table 2:11: Herd-size associated risk factors based on logistic regression

Factor	Outcome	OR	95% CI	p-value
	Proportion spraying			
4-13	17/26 (65.38%)	REF		
14-24	17/22 (77.27%)	1.73	0.51-6.31	0.369
25-46	16/24 (66.67%)	1.05	0.33-3.36	0.924
47-96	21/24 (87.5%)	3.33	0.89-15.08	0.078
97-1000+	24/24 (100%)	26.6	1.45-487.91	0.000*
	Proportion dipping			
4-13	10/26 (38.45%)	REF		
14-24	8/24 (33.33%)	0.92	0.29-2.91	0.881
25-46	10/24 (41.67%)	1.14	0.37-3.48	0.817
47-96	12/24 (50.00%	1.57	0.52-4.82	0.413
97-1000	9/24 (37.50%)	0.96	0.31-2.96	0.944
	Proportion sent away			
4-13	0/26 (0.00%)	REF		
14-24	1/22 (4.55%)	3.69	0.187 – 551	0.990
25-46+	4/24 (16.67%)	11.63	0.59-228.61	0.046*
47-96+	8/24 (33.33%)	27.30	1.48-505.04	0.000*
97-1000+	16/24 (66.67%)	102.88	5.56-1903.08	<0.0001*

Factor	Outcome	OR	95% CI	p-value
	Proportion aware of ECF			
4-13	7/26 (26.92%)	REF		
14-24	7/22 (31.81%)	1.26	0.37-4.29	0.710
25-46	8/24 (33.33%)	1.34	0.41-4.46	0.622
47-96	15/24 (62.50%)	4.24	1.36-14.31	0.013*
97-1000	18/24 (75.00%)	7.40	2.28-27.09	0.001**

[+]Due to complete separation of data, Fishers Exact test was used to calculate the OR

There was no significant association observed between herd size and whether control measures were being used by farmers. A significant association was observed for farmers with the largest herds using the spray method (97-1000, OR = 26.6, p = 0.000). No significant association was made for dipping and herd size. Herd size was significantly associated with whether or not farmers sent their cattle away for periods of time; (25-46, OR = 11.63, p = 0.046; 47-96, OR = 27.3, p = 0.000; 97-1000, OR = 102.88, p = <0.0001).

The proportion of farmers aware of ECF increased significantly with herd size (47-96 cattle, OR = 4.24, P = 0.013, and 97-1000 cattle, OR = 7.40, p = 0.001) (Figure 2.22).

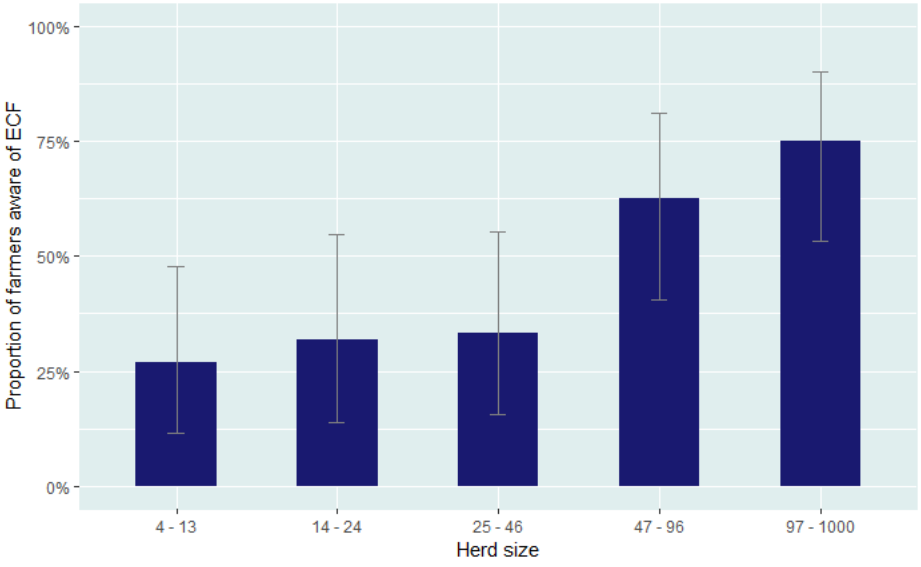

Figure 2.22: Herd size in relation to farmer awareness of ECF

A significant association between awareness of ECF and the proportion spraying their cattle was observed (Table 2:12); the majority of farmers who were aware of ECF were spraying their cattle (89.10%) (OR = 3.19, p = 0.017). There was no significant association between ECF awareness and the proportion of farmers dipping their cattle.

Table 2:12: Logistic regression of farmer awareness of ECF in relation to the proportion of farmers spraying cattle

ECF awareness	Proportion ECF awareness	Proportion spraying	OR	95% CI	p-value
Not aware of ECF	65/120	19/65 (29.23%)	REF		
Aware of ECF	55/120	49/55 (89.10%)	3.19	1.26-9.07	0.017*
		Proportion dipping			
Not aware of ECF	65/120	38/65 (58.46%)	REF		
Aware of ECF	55/120	22/55 (40.00%)	0.94	0.45-1.94	0.864

Herd size was also investigated for association with cattle movement patterns, with the distance travelled for grazing or water (Table 2:13).

Table 2:13: Herd size as a risk factor for cattle movement during wet and dry seasons, based on simple linear regression
The data was log transformed for linear modelling to generate the estimate, standard error, t-value and p-value and was then exponentiated to calculate the multiplicative effect with 95% confidence interval. An offset of 0.05 was added to the model to address distances of zero.

Herd size	Estimate	Standard Error	t value	p-value	Multiplicative estimate with 95% CI
Distance travelled for grazing in wet season					
4-13	-0.443	0.076	-5.826	<0.0001*	0.642 (0.553-0.746)
14-24	0.278	0.107	2.610	0.009*	1.320 (1.07-1.64)
25-46	-0.433	0.095	-4.555	<0.0001*	0.649 (0.538-0.782)
47-96	0.879	0.110	7.985	<0.0001*	2.41 (1.94-2.99)
97-1000	1.019	0.115	8.868	<0.0001*	2.77 (2.21-3.47)

Herd size	Estimate	Standard Error	t value	p-value	Multiplicative estimate with 95% CI
Distance travelled for water in wet season					
4-13	0.003	0.055	0.062	0.951	1.00 (0.900-1.12)
14-24	0.262	0.078	3.355	0.0008*	1.30 (1.11-1.51)
25-46	-0.457	0.069	-6.574	<0.0001*	0.633 (0.553-0.726)
47-96	0.292	0.081	3.626	0.0003*	1.34 (1.14-1.57)
97-1000	0.461	0.084	5.490	<0.0001*	1.59 (1.35-1.87)
Distance travelled for grazing in dry season					
4-13	0.448	0.107	4.199	<0.0001	1.57 (1.27-1.93)
14-24	0.294	0.150	1.960	0.050	1.34 (1.00-1.80)
25-46	-0.265	0.133	-1.987	0.047*	0.77 (0.59-0.99)
47-96	0.143	0.155	0.924	0.356	1.15 (0.85-1.56)
97-1000	1.241	0.161	7.687	<0.0001*	3.46 (2.52-4.75)
Distance travelled for water in dry season					
4-13	1.109	0.063	17.701	<0.0001*	3.03 (2.68-3.43)
14-24	-0.145	0.088	-1.649	0.099	0.865 (0.728-1.03)
25-46	-0.513	0.078	-6.547	<0.0001*	0.599 (0.514-0.698)
47-96	-0.211	0.091	-2.324	0.020*	0.810 (0.678-0.968)
97-1000	-0.419	0.095	-4.426	<0.0001*	0.658 (0.546-0.792)

During the wet season, herds of 14-24 cattle travelled 1.3 times further than the smallest herds (p = 0.009) for grazing. Herds of 25-46 cattle travelled 0.6 times further than the smallest herds (p = <0.0001) and the largest herds (97-1000) travelled 2.7 times further than the smallest herds (p =<0.0001). All herd sizes were significantly associated with travelling further than the smallest herds for water in the wet season. During the dry season, herds of 25-46 cattle travelled 0.7 times further (p = 0.047) for grazing and the largest herds (97-1000) travelled 3.5 times further (p =<0.0001). Herd sizes of 25-46 cattle, 47-96 cattle and 97-1000 cattle were all found to be significantly associated with distance travelled for water during the dry season. Overall, larger herds were significantly associated with travelling (Figure 2.23).

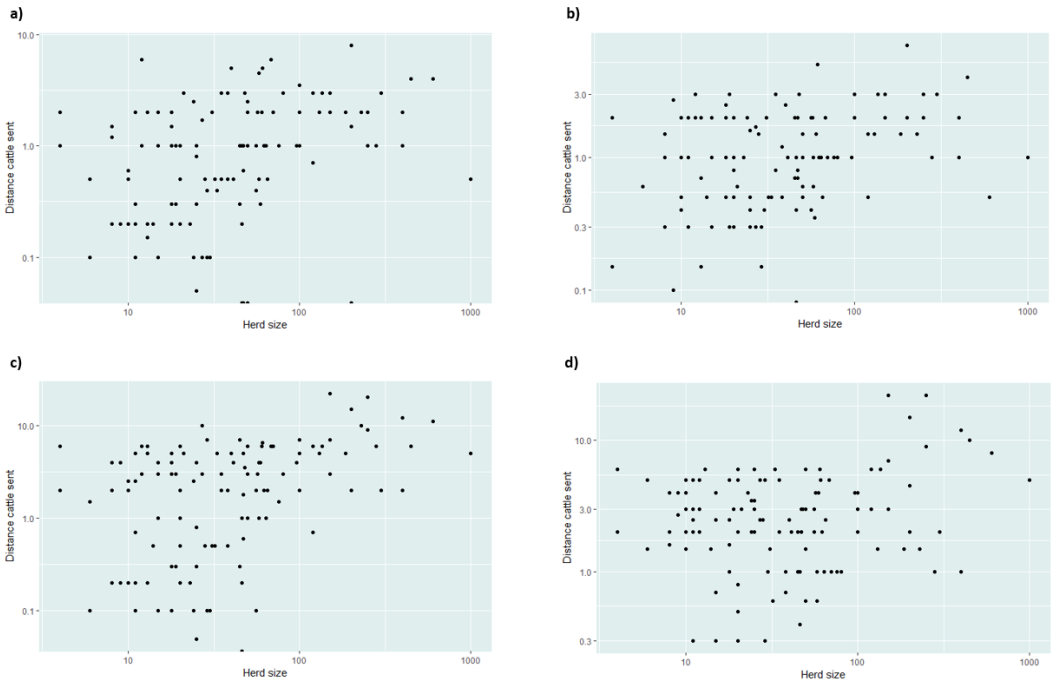

Figure 2.23: Herd-size associations with distances travelled
a) Distances cattle travel for grazing during the wet season, b) Distances cattle travel for water during the wet season, c) Distances cattle travel for grazing during the dry season, d) Distances cattle travel for water during the dry season.

2.6 Discussion

The objectives in this chapter were twofold; firstly to establish the prevalence of *T. parva* in cattle and buffalo in the study area, as well as assessing cow-level risk factors associated with being positive for *T. parva* and secondly to determine vector control practices being used in the area and look at herd-level factors influencing uptake of control. These objectives were achieved and the outcomes will be discussed here.

The overall prevalence of *T. parva* in the 2016 cattle cross-sectional survey was relatively low at 5.07% (CI: 3.70-7.00%), compared to other endemic areas. Prevalence was 100% (CI: 85.00-100.00%) in the buffalo population (22 of 22) in 2011, as was expected due to the endemic nature of *T. parva* in this species (Young et al. 1978).

All the cattle sampled were described as healthy at time of sampling and therefore the *T. parva* positive animals are likely to be recovered carrier animals with very low circulating piroplasms. This stage of infection is characterised by very low parasitaemia (Grootenhuis et al. 1987, Norval et al. 1992), so given the sensitivity and specificity limits of the p104 PCR (discussed in Chapter 3), it is likely that the prevalence is an underestimate of the true proportion of cattle carrying *T. parva.* A study in Kenya detected piroplasms in 37.7% of sampled cattle using standard p104 PCR compared to 42.3% using p104 nPCR, highlighting that even marginal differences in assay sensitivity have great importance in field situations (Odongo et al. 2010).

In the 2016 cross-sectional survey, Robanda had zero prevalence and was significantly different from the other villages, but there were no significant differences between the other individual villages or farms. This may be due to the low level of overall prevalence. Interestingly there was no infection detected in Robanda, despite its proximity to the wildlife areas. An explanation for this may be that there were not many buffalo nearby, as Figure 2.3 demonstrates the distribution of buffalos in the SENAPA. Alternatively this may be due to farmers conducting more vector control due to higher awareness of the risks of disease transmission from wildlife hosts.

In the univariable cow-level risk analysis there was found to be a significant association between older age of cattle and odds of *T. parva* prevalence. This could be explained by the sustained tick challenge these cattle have been exposed to (Magona et al. 2000, Okiria et al. 2002) i.e. a cumulative effect. Additionally, cattle over a year of age tend to be protected from *T. parva* infection and ensuing clinical disease in endemic areas because of acquired immunity after being exposed to infection as calves (Rubaire-Akiiki et al. 2006). Calves are traditionally kept at home and do not graze with adult cattle in order to limit their exposure to ticks (Muhanguzi 2010a, Muhanguzi 2010b). Calves associate with their dams only during milking, when calves can stimulate the let-down of milk, and thus calves remain susceptible to TBDs, particularly so at around six months of age when maternal antibodies wane and at this time

they can seroconvert if exposed to *T. parva* and progress to developing clinical disease (Muhanguzi et al. 2014a). Although calves under 6 months were not included in this survey, the odds of infection was 3.47 (CI: 1.39-9.39%) times higher in cattle aged over 4.5 years, compared to those aged between 6 months and 1.5 years.

The 2016 survey found very low counts of ticks on cattle, particularly ear ticks, which include *R. appendiculatus*, which were observed on only 3.66% of cattle. The time of sampling, in July and August 2016, followed an unseasonably wet period. Climate is very important in tick distribution and changes in rainfall can be highly influential on tick activity (Kimaro 2013, Olwoch et al. 2008). Low tick counts can be due to good vector control management (Kazungu et al. 2015a). There is the risk where intensive and effective tick control measures are being used that endemic instability is created and more animals succumb to ECF (Kivaria et al. 2012). As part of the sampling survey, pilot work was carried out involving surveys for ticks in the environment. Using both blanket drag and sweep net techniques, vegetation transects were sampled in protected areas as well as farming areas around the SENAPA. However, low numbers of ticks caught precluded any analysis and so this work was not continued. Preliminary data are included in Appendix D.

The questionnaire established that the majority of farmers were using acaricides as vector control. All of the acaricide products being used advise fortnightly application. The majority of responding farmers reported using acaricides weekly, with some farmers treating every four days. In order for treatments to be effective in controlling ticks, application of acaricides must adequately interrupt the life cycle of the targeted tick species as well as have residual effectiveness. In the case of *R. appendiculatus* it is reported that it can be necessary to treat cattle on a weekly basis and if disease transmission prevention is required the treatment interval may be as often as every 3-5 days (FAO 1984). Certainly in regions where ECF is endemic and where exotic cattle breeds are present, it is reported that high losses can only be prevented with weekly acaricide application (Seifert 1996). A study by Kerario et al (2018)

in Southern Highlands, Tanzania, found that most farmers were spraying or dipping their animals every 2 weeks. The application of acaricide by dipping animals once every two weeks in pastoral and ago-pastoral communities in eastern Tanzania had been reported to be economical as it reduced the costs of acaricide and limited the losses caused by the death of the animals due to TBDs (Mbassa G K 2009). The frequent application of acaricide may reduce transmission of *T. parva* but immune cattle will continue to be carriers of *T. parva* and acaricides will not eradicate the disease (Kariuki et al. 1995).

Despite their frequent use, only a small proportion of acaricides for spraying were being prepared correctly with most cattle being underdosed; 100% farmers were dosing with Tantix incorrectly, 81.82% were incorrectly dosing with Cybadip, 58.06% were dosing incorrectly with Paranex and 56.52% were dosing with Albadip incorrectly. This is consistent with a questionnaire survey by Swai et al. (2009) which reported farmers in Tanga, north-east Tanzania, using acaricides at incorrect dosage and frequency and another survey by Swai et al. (2005a) in south-east, Tanzania, found no farmers to be using correct concentrations. Ogden et al. (2005) and Swai et al. (2005) found incorrect dilutions of acaricides on farms in surveys in southern highland and north-east Tanzania. Incorrect use includes overdosing which can result in toxicity to animals (Addah 2009) and be harmful to farmers and the environment as well as creating unnecessary additional cost (Swai 2005a). If the improper use of acaricide is far-reaching and has been in effect for many years, it raises concerns about sustainability, with tick resistance a real concern when underdosing cattle. Studies in Uganda by Vudriko et al. (2016) using the larval packet test, demonstrated high levels of resistance to synthetic pyrethroids in *R. appendiculatus*, even when doubling the concentration of acaricide. According to a report by the Food and Agriculture Organisation of the United Nations (FAO), resistance to synthetic pyrethroids and organophosphate acaricides is “widespread”, with some resistance to amitraz also reported (FAO 2004). The Veterinary Parasite Resistance Group (VPRG) was established in 1995 to advise the FAO how best to manage resistance, and they describe resistance to pyrethroids appearing in the 1980s (Graf et al.

2004). Due to their similar modes of action, pyrethroids showed cross-resistance to the organochlorine Dichlorodiphenyltrichloroethane (DDT), with genuine pyrethroid resistance developing and rapidly spreading in the 1990s (Graf et al. 2004). Unpublished studies by Lynen et al. and the Tropical Pesticide Research Institute (TPRI), Tanzania, in 2002-2003 found very high levels of resistance to synthetic pyrethroids and given they had only been available on the Tanzanian market for a few years, it is thought that the resistance developed quickly, likely due to wide-spread aerial spraying of DDT for crop protection and malaria control (personal communication, Dr Lieve Lynen).

The survey highlighted a basic lack of instruction for preparation and application of acaricide for farmers to follow; in some instances the instructions provided did not state the required volume per head or the frequency of application and often the directions were only provided in English which is not the native language of Tanzanian farmers and many of whom have low level literacy. Owners of larger herds were found to be significantly more likely to be aware of ECF and owners who were more aware were more likely to spray their cattle. Although education levels were not measured, it is likely that the farmers of smaller herds in particular may have less access to information, making them more vulnerable. Adherence to manufacturer guidelines is a necessity in order to achieve the most effective tick control (George et al. 2004). A host of operational factors in the application of acaricides can also reduce the effectiveness of tick control, for example inadequate replenishment of dipping solution in dip tanks, incomplete stirring of dip tank solution and insufficient wetting of cattle with spray (George et al. 2004). It was ascertained that five of the eight sampled villages in the 2016 cross-sectional survey had functioning dip tanks (Bwitengi, Park Nyigoti, Makundusi, Robanda and Singisi). Nyiberekera, Bisarara and Tamkeri did not have dip tanks (Personal communication, Emmanuel Sindoya). A survey by Mwaseba and Kigoda (2017) described limited access to dipping in Serengeti District; a total of 91 villages had only 36 dip tanks and of those only 21 were operational. In other

studies in Tanzania, farmers reported lack of dips to be a constraint to cattle production, with many dips not functioning (Chenyambuga 2010).

It was established that all farmers in the study were using a single drug class of acaricide - cypermethrins - without any drug rotation, where 'rotation' is the strategic alteration of treatments over time with chemicals with varied modes of action (Riddles and Nolan 1987). This has serious implications in the potential development of tick resistance to acaricides. Emergence of tick resistance has been reported in Tanzania (George et al. 2004). Rotation policies could extend the lifespan for individual acaricides and mixing acaricide classes could delay the development of resistance (Dolan 1999). However, during discussions with livestock policy stakeholders, it emerged that it is national policy in areas where ticks and tsetse co-occur for acaricides that are effective against both vectors to be used – indeed there is a government subsidy of synthetic pyrethroids in such areas (Personal communication, Joyce Daffa, Tsetse Control Division). This provides an explanation for the use of a single compound in the area, but does not remove concerns about sustainability of acaricide use in the absence of any rotation.

None of the farmers in the questionnaire survey were using the ITM vaccine. This was predominantly due to lack of awareness of the vaccine. One farmer reported the cost to be prohibitive. The vaccine costs between 15,000 and 20,000 TZS per cow and 11,000-12,000 TZS per calf (personal communication, Livestock Officer Emmanuel Sindoya). Vaccine uptake has a strong association with affluence and this necessity for economic security precludes poorer farmers from accessing the vaccine for their cattle (Homewood et al. 2006). The vaccine is used quite widely in other parts of Tanzania (Di Giulio et al. 2009, Lynen et al. 2012, Martins et al. 2010)

Farmer knowledge about ticks and ECF varied. Farmers from both cross-sectional and longitudinal herds had good awareness of ticks but cross-sectional farmers were less aware of the diseases transmitted and clinical signs associated. Longitudinal farmers had significantly larger herds than the cross-sectional herds sampled in 2016. There was a significant association

found between larger herd size and better awareness of ECF. As mentioned, the longitudinal farmers tended to be knowledgeable and proactive with regard to herd health, exemplified by their initial volunteering to report FMDV outbreaks in original recruitment in 2013. The larger herd sizes suggested greater wealth and perhaps better education (although this was not assessed) and so may be expected to show greater knowledge compared to the cross-sectional farmers.

There were several limitations of the study. There can be a lack of reliability in the use of a design effect in calculating sample size where there is clustering (Thrusfield 2005). The design effect of two used was based on published values for surveys of vector-borne diseases in sub-Saharan Africa (Otte and Gumm 1997), and in this instance the sample size far exceeded the required calculated sample size. For some questions, the questionnaire response rate was poor. In Section 2.5.10, for example, farmers were asked if they used prevention methods for ECF and only 38 of 120 farmers responded. Given that the farmers had already reported spraying and dipping as prevention methods for ticks it may have been assumed that they were spraying and dipping as prevention for ECF. It is also important to consider whether farmers reported accurate answers; as the questionnaire enumerator was a local livestock officer known to the farmers, they may have felt that they should report the ‘correct’ answer which may have differed from the ‘real’ answer. Although in-person interviews can have interviewer bias, this type of interview generally has less missing data compared to postal questionnaires (Smeeth et al. 2001) and so there is a trade-off in how the questionnaire is completed depending on whether or not an enumerator is involved.

In the herd-level risk analysis there was a significant association observed between the larger herd sizes and the uptake of spraying. No such association was made for dipping. Overall most farmers were spraying (79.16%, CI: 71-85%) with 40.83% (CI: 32-49%) dipping, and 20.83% (14-29%) of farmers were using a combination of both spraying and dipping. No one reported using pour-on products or applying acaricide with brushes as is done in other areas.

Hand-spray application of acaricide is considered the most popular method small-scale farmers use (Mugisha et al. 2005). FAO guidelines (1984) suggest the use of hand spraying should only be used in herds of less than 10 cattle due to the laborious and potentially less thorough nature of this method compared to dipping. Hand spraying, however, can be highly ineffective if the preparation and application of the acaricide is insufficient; in an attempt to reduce the cost of the formulation farmers may use inadequate amounts (De Meneghi et al. 2016).

A significant herd-size association was made with larger herds being sent away for periods of time as well as larger herds travelling further for grazing and water – larger herds have greater requirements for grazing and water sources and so are forced to travel further. Livestock movement is a risk factor for the spread of disease (Macpherson 1995) and grazing practices are a component of general animal husbandry that affect the risk of disease transmission due to contact between herds (Bronsvoort et al. 2004, Schoonman and Swai 2010). The necessity to move livestock for water, especially during the dry season, can result in multiple cattle herds and wildlife sharing water sources, allowing for pathogen transmission (Bouslikhane 2015, Katale et al. 2013). Despite often moving their cattle long distances, all participants of the questionnaire reported bringing their cattle back to the boma nightly. However, communal and transhumant grazing practices (seasonal movement patterns) are common in arid areas of Tanzania and these practices can increase the contact rate between herds and therefore the transmission of disease (Hutchings and Harris 1997). In the case of ECF and other TBDs, cattle movement through areas with abundant tick populations could increase exposure risk. Control measures used to mitigate disease transmission while travelling will depend on farmer awareness of risk.

In addition to the data collected in the surveys described, two workshops were held with farmers, village leaders and livestock extension officers in the study area in February 2017. The workshops comprised focus group discussions to a) provide some validation of questionnaire responses and explore more

qualitative responses such as drivers for use of different control methods and how vector control has changed over the last five years, and b) feedback some preliminary data to encourage farmer engagement with the project. This provided some very useful insights into farmer and veterinary perspectives on vector control. Workshops confirmed that the vector control measures in the study area were predominantly farmer-led. There was not a clear concept of if and how practices had changed over time, although some farmers did report that they had started using vector control only in the past few years. Farmers expressed concern at the effectiveness of products they were using, stating the necessity to treat with increased frequency due to poor product efficacy. This highlights the need for consideration of the development of tick resistance as well as the possibility of substandard or counterfeit products. Drug-resistance and treatment failure due to poor manufacturing are reported consequences of substandard product use (Shakoor et al. 1997).

It is likely that the low prevalence of *T. parva* observed in the 2016 cross-sectional survey is due to the high frequency use of acaricides in the study area and thus the low tick counts. It is unclear when farmers began using these vector control products but if they are a relatively recent management practice they could explain the lower prevalence of *T. parva* compared to other hyperendemic areas.

Despite farmers often not following recommended guidelines, the lack of ticks found on animals and low prevalence of infection suggest that the control of ticks and East Coast fever in the study currently is effective. The sustainability of such an acaricide-dependent control strategy, however, remains uncertain. Farmers in the study area are concerned about ticks and are using continuous acaricide application throughout the year. All of the farmers were using only cypermethrin drugs and most of the farmers were found to be underdosing their cattle. None of the farmers were using the ITM vaccine. Awareness of ECF was varied, with farmers of larger herds being significantly more aware. Despite government policies regarding acaricide use, workshops demonstrated that the vector control in the area was farmer-led. There is

scope, therefore, to engage with farmers and work to improve disease awareness and acaricide use in the study area.

Chapter 3: Establishing a genotyping pipeline to assess antigenic diversity

3.1 Introduction

The Infection and Treatment Method (ITM) of vaccination can protect cattle from East Coast fever (ECF) (Radley 1981). The ITM utilises a mixture of three *Theileria parva* stocks (uncloned isolates) – Muguga, Serengeti-transformed and Kiambu 5 – known as the Muguga cocktail vaccine (Radley 1975a, Radley 1975b, Radley 1975c). Previous analyses of the Muguga cocktail components identified limited diversity and significant similarity between the Muguga and Serengeti-transformed stocks (Bishop et al. 2001, Oura et al. 2007, Oura et al. 2004a). More recent studies by Norling et al. (2015) and Hemmink et al. (2016) have identified remarkable similarities between these two stocks, with only 420 non-synonymous single nucleotide polymorphisms (SNPs) between the two sequenced genomes – while the third stock, Kiambu 5, is significantly more divergent, with almost 40,000 SNPs relative to the reference Muguga genome.

Studies in Tanzania have shown the vaccine to confer protection from cattle-derived *T. parva* (Komba 1991, Martins et al. 2010, Uilenberg et al. 1977, Uilenberg et al. 1976). It is known that there is far greater genotypic diversity in buffalo-derived *T. parva* parasites compared to the population maintained in cattle (Bishop et al. 1994a) and current evidence indicates that there is only a limited subset of the *T. parva* gene pool in buffalo that circulates in cattle (Oura et al. 2011a, Pelle et al. 2011). A study by Sitt et al. (2015) showed that cattle vaccinated with the Muguga cocktail before introduction onto a site previously grazed only by buffalo were not protected against developing the clinical disease associated with ECF. The vaccine does, however, provide protection in some areas that are co-grazed by both cattle and buffalo (Di Giulio et al. 2009). The genotypic composition and origins of the parasites to which grazing cattle are exposed have not been well studied. There is a need for better understanding of genetic and antigenic diversity of field populations of *T. parva* in order to improve use of the current vaccine and, in time, aid progress to developing improved methods of vaccination.

Immunity of cattle to *T. parva* is T cell-mediated (Morrison et al. 1987), targeting schizont-infected cells (Daubenberger et al. 1999). There is evidence that CD8 T cells play a role in protection, with clearance of *T. parva* infection coinciding with the appearance of CD8 T cells (Morrison et al. 1981, Pearson et al. 1979) as well as adoptive transfer studies of CD8 T cells between immune and naïve calves showing protection (Emery 1981, McKeever et al. 1994). The identification of several *T. parva* schizont antigens, and epitopes within those antigens, recognised by CD8 T cells from immune cattle (Graham et al. 2006) allowed for a series of studies to examine CD8 T cell antigen specificity, and thus strain specificity of immunity (Morrison et al. 2015). There is a significant inter-relationship between CD8 T cell specificity and distinct parasite strains and thus a correlation with the ability to withstand heterologous challenge (Taracha et al. 1995b). The molecular basis of this specificity is not fully understood but one important determinant is polymorphism in the antigens recognised by the CD8 T cells (MacHugh et al. 2009). Antigens have been identified by two approaches; first, cDNAs for candidate genes were selected by examining the *T. parva* genome and the expressed products then screened *in vitro* for recognition by CD8 T cells and second, *T. parva* specific CD8 T cell lines were used to screen a parasite schizont cDNA library (Graham et al. 2007, Graham et al. 2006). Both approaches involved co-expression of the parasite genes with relevant class I MHC genes in COS-7 cells. These approaches identified eight CD8 T cell target antigens – named Tp1-Tp8 (Akoolo et al. 2008, Graham et al. 2007, Graham et al. 2006). Tp1 encodes a 543 amino acid protein (Pelle et al. 2011) and although its role has not been formally identified (Connelley et al. 2011, MacHugh et al. 2009), Tp1 is known to be an immunodominant antigen that induces a cytotoxic CD8 T cell response (MacHugh et al. 2009). Previous studies of Tp1 have shown extensive sequence diversity, especially in buffalo-derived *T. parva* parasites (Pelle et al. 2011). Tp2 encodes a protein of 174 amino acids and has been shown to be highly polymorphic (Hemmink et al. 2016, Pelle et al. 2011). A study by Hemmink (2014) also found allelic variation in Tp4, Tp5 and Tp6 at the nucleotide level, but almost all SNPs were

synonymous resulting in conservation at the amino acid level. Additional studies identified three more antigens – Tp9, Tp10 and Tp12 – by screening with CD8 T cell lines from highly susceptible immunised Holstein cattle (MacHugh, N.D., Graham, S. and Morrison, W.I., unpublished). Tp9 has demonstrated marked sequence diversity (Hemmink et al. 2016).

CD4 T cell involvement in immunity to *T. parva* has been minimally investigated, but strong CD4 T cell responses to infected cells in immune animals were observed in early studies (Baldwin et al. 1987, Brown et al. 1989b, Goddeeris and Morrison 1988). CD4 T cells recognise antigens presented by MHC class II on infected cells (Brown et al. 1989b). Studies carried out in the 1980s showed CD4 T cell clones recognised 'subcellular fractions' of parasitized cells, identifying soluble and schizont membrane-associated antigens (Baldwin et al. 1992, Brown et al. 1990, Brown et al. 1989b, Grab et al. 1992). Taracha et al. (1997) showed that optimal *in vitro* activation of CD8 T cells from immune cattle requires CD4 T cells, suggesting an important role for CD4 T cells in the immune response. Ongoing investigation of the specificity of CD4 T cells, utilising the screening of parasite cDNA and peptide libraries (Morrison et al. 2015) and unpublished studies (Morrison, W.I.), has identified several CD4 T cell antigens, including N10 (Tp14), N43, A14 and N60 (Tp16).

Unlike the antigens recognised by T cells, the p67 protein is a sporozoite surface antigen, recognised by neutralizing antibodies (Graham et al. 2007a) and is considered a potential protective *T. parva* antigen (Dobbelaere et al. 1985a, Tebaldi et al. 2017) and possible vaccine candidate. The expression of p67 is restricted to the sporozoite stage (Dobbelaere et al. 1985b). Musoke et al. (1992) showed that antibodies against p67 inhibit parasite entry into cells and are able to neutralise infection *in vitro*. Studies by Nene (1996) and Sibeko (2010) found one allele of p67 in cattle-derived *T. parva*, but found four alleles, with amino acid differences, in buffalo-derived *T. parva*.

Theileria parasites infect domestic and wild ruminant species and are categorised as either schizont-transforming or non-transforming (Sivakumar et

al. 2014). Schizont-transforming (pathogenic) *Theileria* parasites are contained in the *T. taurotragi* clade (Sivakumar et al. 2014), which includes *T. parva*, *T. annulata*, *T. taurotragi* and *T. lestoquardi* (Dobbelaere and Kuenzi 2004, Sugimoto and Fujisaki 2002); it is the uncontrolled proliferation of schizont-infected cells that causes the pathology associated with East Coast fever in cattle (Bishop et al. 2004, McKeever 2009). The benign parasites *T.* sp. (buffalo) and *T.* sp. (bougasvlei) are also part of the *T. taurotragi* clade (Mans et al. 2015, Mans et al. 2011, Pienaar et al. 2014) but although largely considered apathogenic (Allsopp et al. 1993, Zweygarth et al. 2009), the piroplasm stage can cause anaemia (Sivakumar et al. 2014).

Being from the same clade, the transforming species *Theileria* sp. (buffalo) is genetically closely related to *T. parva* (Mans et al. 2011, Pienaar et al. 2011, Zweygarth et al. 2009). *T.* sp. (buffalo) was initially shown only to be present in buffalo (Allsopp et al. 1999, Mans et al. 2011, Oura et al. 2004b, Oura et al. 2011b, Pienaar et al. 2011) but has recently been identified in acute infections in cattle, along with *T. parva* (Hemmink 2014), on a ranch in the Rift Valley District of Kenya, where buffalo are known to graze. Due to a high degree of similarity of *T. parva* and *T.* sp. (buffalo), the presence of both in mixed infections can complicate diagnosis by PCR.

The work presented in this methodological chapter describes the establishment of a genotyping pipeline, which will be applied in Chapter 4 to full-length PacBio sequencing data in order to analyse antigenic diversity of field populations of *T. parva* at the livestock/ wildlife interface, where cattle and buffalo can interact and parasite populations can mix. In order to obtain full-length or near full-length sequences, single-molecule real-time (SMRT) sequencing, third-generation sequencing (TGS) technology developed by PacBio (Rhoads and Au 2015), was utilised. Real-time sequencing involves sequencing of single DNA molecules (Schadt et al. 2010). Single-molecule sequencing (SMS) has increased the rate of sequencing along with increasing the lengths of reads generated (Rhoads and Au 2015).

The workflow includes establishing a robust diagnostic assay to detect *T. parva* in field samples, and in particular one that clearly differentiates between *T. parva* and *Theileria* sp. (buffalo), the design and trial of PCR primers to amplify antigen genes, optimisation of sample preparation, and the creation of a bioinformatics workflow for sequence analysis downstream, with the overall aim being to create a pipeline that will allow analysis of the antigenic diversity in field parasite populations and to compare diversity in and between cattle and buffalo.

3.2 Objectives

1. Design a pipeline for amplifying and analysing polymorphic sequences
    - Identify the most robust diagnostic assay to identify *T. parva* in field samples
    - Design primers and PCR assays to amplify polymorphic antigen genes across diverse *T. parva* cloned isolates – to be applied to positive cattle and buffalo samples from Ol Pejeta game conservancy, Kenya
2. Generate pilot data from cattle and buffalo in order to validate primers and PCR assays and the application of PacBio long-read sequencing technology

A previous study by Hemmink (2014) provided sequence data for six parasite genes from eight buffalo samples, by Roche 454 sequencing of PCR amplicons. These six genes – Tp1, Tp2, Tp4, Tp5, Tp6 and Tp10 – are known to encode *T. parva* antigens recognised by CD8 T cells (Graham et al. 2007, Graham et al. 2006). The 454 sequencing approach limited the size of amplicons that could be examined (<450 base pairs), and therefore the sequence reads only captured part of the genes. The sequence data for the six genes identified multiple alleles of all genes within individual animals. Tp1

and Tp2, however, showed extensive polymorphism at the amino acid level (Pelle et al. 2011) and separate studies have shown Tp9 also to be highly diverse (Hemmink et al. 2016) so these three genes were of particular interest. Tp12 and the CD4 T cell antigens - N10 (Tp14), N43, A14 and N60 (Tp16) – were selected on the basis of unpublished studies (Morrison, W.I.), which showed varying degrees of diversity in cattle and buffalo. p67 was selected on the basis of it being a vaccine candidate antigen (Nene et al. 1996, Tebaldi et al. 2017).

3.3 Materials and Methods

3.3.1 Cell culture

In order to design and validate PCR primers, it was necessary to grow cultures of *T. parva* Muguga, the genome reference parasite, in order to provide positive control genomic DNA of a validated clonal lineage. Cryovials of cell line 641 *T. parva* Muguga-infected cells in cryoprotectant were removed from liquid nitrogen and warmed in a water bath at 37°C until thawed. The cells were then transferred to a 15 ml tube and slowly diluted in 6 ml culture medium [RPMI-1640 (Sigma-Aldrich®) medium supplemented with 50 ml FCS, 5 ml Penicillin-Streptomycin-Glutamine (ThermoFisher Scientific) and 500 µl β-mercaptoethanol]. The sample was centrifuged at 1500 x *g* for five minutes and the supernatant was discarded before resuspending the cells in a further 6 ml culture medium. The cell suspension was checked for viability and transferred to a 75 ml flask and incubated at 37°C and 5% CO_2. The cells were checked microscopically for growth and viability and sub-cultured every 2-3 days to expand the cell numbers.

3.3.2 DNA extraction

GeneJET DNA Purification Protocol (Thermo Fisher Scientific) was used to extract DNA from cultured *T. parva* Muguga-infected cells. Suspended cells infected with 641 *T. parva* Muguga were counted using a haemocytometer,

and $5x10^6$ cells added to a centrifuge tube. Centrifugation was carried out at 250 x *g* for five minutes to pellet the cells. The supernatant was discarded and the cells washed with Phosphate Buffered Saline (PBS) to remove any residual medium and centrifugation was repeated, again discarding the supernatant. Cells were resuspended in 200 µl of Tris-EDTA (TE) buffer. 200 µl Lysis Solution and 20 µl Proteinase K Solution were added to the cell pellet and then mixed by vortexing. The sample was incubated at 56°C for 10 minutes, while vortexing intermittently. 20 µl RNase A Solution was added and mixed before incubating at room temperature for 10 minutes. 400 µl 50% ethanol was added and mixed before transferring the newly prepared lysate to a GeneJET Genomic DNA Purification Column inserted in a collection tube. The column was centrifuged for one minute at 6,000 x *g*, the collection tube discarded and the column placed into a new 2 ml collection tube. 500 µl Wash Buffer I (with ethanol added) was added and the tube centrifuged for one minute at 8,000 x *g.* The flow-through was discarded and the purification column replaced into the collection tube. 500 µl Wash Buffer II (with ethanol added) was added to the Purification column and centrifuged for three minutes at maximum speed (≥ 12,000 x *g*) and if residual solution was seen in the column, the collection tube was emptied and the tube re-spun for a further one minute at maximum speed. The collection tube was discarded and the Purification column transferred to a sterile 1.5 ml Eppendorf tube. 200 µl Elution Buffer was added to the Purification column and incubated at room temperature for two minutes before centrifuging at 8,000 x *g* for one minute to elute the genomic DNA. In order to yield maximum DNA, the elution step was repeated with an additional 200 µl Elution Buffer. The purification column was discarded and the purified DNA stored at -20°C

3.3.3 PCR

For the diagnostic assay targeting p104 the PCR mix consisted of 12.5 µl Quick-Load Taq 2X Master Mix (New England Biolabs), 1 µl of each primer (10µM), 10 µl nuclease-free water and 1 µl of DNA template (total = 25 µl

reaction). For initial working up or primers, a conventional single round PCR was used; for subsequent development, a nested PCR was used. For the second round template the first round product was diluted 1:100 in dH_20. Primer sequences and cycling conditions are shown in Table 3:1.

PCR amplification of the target antigen genes for sequencing used a high fidelity DNA polymerase because of the > 100 fold lower error rate compared to *Taq* DNA polymerase, making it suitable for generation of accurate sequence data from nested PCR amplicons. PCR mix consisted of 5 µl Q5 High-Fidelity Reaction Buffer (New England Biolabs), 0.5 µl dNTPs (Bioline, Meridian Life Science Company), 1.25 µl of each primer (10µM), 0.25 µl Q5 High Fidelity DNA Polymerase (New England Biolabs), 15.75 µl nuclease-free water and 1 µl of DNA template (total = 25 µl reaction). For the second round template the first round product was diluted 1:50 in DH_20. Final primer sequences and PCR conditions are shown in Table 3:5.

All PCR reactions were carried out in a MJ Research PTC-200 DNA Engine thermal cycler, and the nPCR products were visualised by UV trans-illumination in a 1.5% agarose gel containing GelRed (Biotium) after electrophoresis. A 100-bp or 1kb DNA ladder was used depending on expected product size (Promega). Positive controls (DNA of *T. parva* Muguga) and negative controls (no DNA template) were included in each PCR.

3.3.4 PCR purification

PCR products were purified using the QIA quick PCR Purification Kit Protocol (Qiagen). For every one volume of PCR product, five volumes of Buffer PB was added and both were mixed. Buffer PB contains a pH indicator and, the colour of the mixture was checked to be yellow, indicating optimal pH for DNA adsorption (pH ≤ 7.5). The mixed sample was placed in a QIAquick spin column and centrifuged for sixty seconds at 17,900 x *g*. The flow-through was discarded and the column placed back into its collection tube. The sample was next washed with 0.75 ml Buffer PE and centrifuged again for sixty

seconds at 17,900 x *g*, discarding the flow-through and repeating centrifugation once again for an additional sixty seconds to ensure residual ethanol was removed. The column was then placed in a clean microcentrifuge tube and 30 µl nuclease-free water added to the column membrane to elute the DNA; to maximise DNA concentration, the column was left to stand for sixty seconds before centrifuging for sixty seconds at 17,900 x *g*. The purified DNA was then stored at 4°C or -20°C if longer storage was required.

DNA purity was assessed using NanoDrop ND-1000 spectrophotometer (Thermo Fisher Scientific), which assessed the 260/280 and 260/230 measurement ratios for any impurities. Some of the 260/230 ratios were slightly lower than required for sequencing downstream so a further clean-up procedure was indicated. Ethanol precipitation was carried out to further purify the genomic DNA in these instances; cold ethanol (100%) was added at two times the volume of DNA before adding 3M sodium acetate at 10% of the original volume. The sample was mixed by inverting several times before chilling at -80°C for thirty minutes to aid precipitation. The DNA was pelleted by centrifuging at 4°C at 17,900 x *g* for twenty minutes, making sure to note the orientation of the tubes in the centrifuge so as to identify the expected location of the pellet. The supernatant was removed whilst ensuring the pellet was not disturbed. The pellet was washed with 200 µl of cold ethanol (75%) and mixed by inverting to ensure washing was thorough. The pellet was centrifuged at 4°C at 17,900 x *g* for five minutes, the supernatant removed before centrifuging once again for one minute to remove residual ethanol. The pellet was air-dried in a fume hood for approximately ten minutes before resuspending the DNA in nuclease-free water and incubating at 37°C for 30 minutes to aid resuspension. The volume of water added to resuspend was dependent on the starting concentration of DNA and the desired end concentration.

3.4 Results and method development

3.4.1 Diagnostic assay

In order to detect *T. parva* infection within field samples, a robust diagnostic PCR assay was required. This assay needed to be sensitive because in field samples the number of parasites in peripheral blood is often very low (particularly in carrier rather than clinically ill animals), therefore a nested PCR (nPCR) assay was indicated. As well as being sensitive, the diagnostic assay was also required to be specific as it was crucial that *T. parva* could be discriminated from the multiple other *Theileria* species that have been identified in East African cattle in the field – particularly in a context where cattle are co-grazing with wildlife, which harbour many species of *Theileria* also infective to cattle (Kariuki et al. 2012, Lawrence et al. 1983). Several *Theileria* species are commonly found in African buffalo, including *T. parva*, *T. buffeli*, *T. mutans, T. velifera* and *T*. sp. (buffalo) (Allsopp et al. 1999, Hemmink et al. 2018, Oura et al. 2011a, Young et al. 1978). It was previously thought that *T. buffeli* and *T.* sp. (buffalo) only infected buffalo but relatively recently Oura et al. (2011b) found *T. buffeli* in cattle, in Lake Mburo National Park, Uganda and as mentioned previously, *T.* sp. (buffalo) has also recently been found in cattle (Hemmink 2014). *T. mutans* and *T. velifera* are considered usually non-pathogenic (Oura et al. 2011a) although *T. mutans* can be fatal if causing severe anaemia in susceptible cattle (Snodgrass et al. 1972). The presence of alternative non-pathogenic *Theileria* species in mixed infections can complicate the specific diagnosis of *T. parva* (Zweygarth et al. 2009). This can be challenging at the molecular level due to significant sequence similarity (Bishop et al. 1994b).

An existing assay targets the amplification of the *T. parva*-specific p104 gene, with primers designed from conserved regions of this gene, generating a 496 bp PCR product (Skilton et al. 2002) (Table 3:1). To further enhance the sensitivity of the assay Odongo et al. (2010) developed a nested p104 assay, amplifying a 277 bp internal fragment. Serological methods such as enzyme-linked immunosorbent assay (ELISA) have been developed to detect specific

antibodies which reflect previous exposure to infection, but the p104 PCR assay identifies current infections and thus provides a sensitive means of detecting infection, estimated to detect 0.4 parasites/µl (equating to a blood parasitaemia of 9.2 x 10^{-6}%) (Odongo et al. 2010) with a reported specificity of 100% (Skilton et al. 2002) and is comparatively more sensitive than other existing assays (Gubbels et al. 1999, Oura et al. 2004b, Skilton et al. 2002). Nested PCR is intended to increase specificity (Nicolaiewsky et al. 2001, Odongo et al. 2010, Schnittger et al. 2004, Ueti et al. 2012). This nPCR assay is considered both specific and sensitive for detecting low levels of *T. parva* infections in field samples (Odongo et al. 2010, Skilton et al. 2002).

Table 3:1: p104 primers and cycling conditions used in nested PCR

	Primers	Sequence	Cycling conditions
Round 1	IL3231	ATTTAAGGAACCTGACGTGACTGC	94°C for 1 min, 40 cycles (94°C for 1 min, 60°C for 1 min, 72°C for 1 min), 72°C for 9 min
	IL755	TAAGATGCCGACTATTAATGACACC	
Round 2	p104_f2_Od	GGCCAAGGTCTCCTTCAGATTACG	94°C for 1 min, 30 cycles (94°C for 1 min, 55°C for 1 min, 72°C for 1 min), 72°C for 9 min
	p104_r2_Od	TGGGTGTGTTTCCTCGTCATCTGC	

(Odongo et al. 2010, Skilton et al. 2002)

A panel of DNA samples (Table 3:2) from cloned parasitized cell lines was compiled in order to test the specificity of the p104 primers – the number and diversity of isolates allowed robust assessment of specificity, and in particular enabled a stringent analysis of cross-amplification (or not) between *T. parva* and *T*. sp. (buffalo). This panel included DNA from cloned isolates of eight *Theileria* species, namely *T. annulata*, *T. buffeli*, *T. taurotragi*, and five different *T*. sp. (buffalo) clones, and uninfected *Rhipicephalus appendiculatus* tick DNA as a negative control. Primer specificity was also trialled on multiple cloned *T. parva* isolates derived from African buffalo from the Masai Mara District and Laikipia District, Kenya (Conrad et al. 1987, Hemmink 2014) (n = 14) as well as cloned isolates originating from cattle used in a vaccine trial in Rift Valley District, Kenya (Hemmink 2014, Pelle et al. 2011, Young et al. 1992) (n=8); the latter study area was grazed by buffalo so these isolates represent buffalo-

associated *T. parva* clones. By including buffalo-associated *T. parva* isolates in the panel, primers could therefore be tested across multiple allele variants, as these parasites are known to be more genetically diverse compared to cattle-derived *T. parva* parasites (Bishop et al. 1994a, Collins and Allsopp 1999, Conrad et al. 1989, Geysen et al. 2004, Hemmink 2014, Pelle et al. 2011).

DNA aliquots of the test panel (provided by Professor Ivan Morrison, Roslin Institute, Edinburgh) were standardised to 60 ng/µl. Uninfected *R. appendiculatus* tick DNA was acquired using a tick from the colony at The Roslin Institute (colony originally established at ILRI, Kenya), and by tick homogenisation and subsequent DNA extraction. All isolates in the panel had previously been sequenced at the 18S rRNA locus to attest their correct identification as *T. parva* (Gubbels et al. 1999, Hemmink 2014, Oura et al. 2004b).

Table 3:2: DNA panel used to test primer specificity

DNA Sample ID	Country of Origin	Sample type
T. parva (Muguga)	Kenya	Reference genome
T. annulata	Turkey	Non-pathogenic/alternative *Theileria* species
T. buffeli	Kenya	
T. taurotragi	Kenya	
T. sp. (buffalo) 6834 clone 10	Kenya	
T. sp. (buffalo) 6998 clone 10	Kenya	
T. sp. (buffalo) 6834 clone 5	Kenya	
T. sp. (buffalo) 6998 clone 2	Kenya	
T. sp. (buffalo) 6998 clone 4	Kenya	
Rhipicephalus appendiculatus DNA	Kenya (ILRI)*	Negative control
Buffalo clone M3.3	Kenya	Buffalo-derived *T. parva* clones
Buffalo clone M3.6	Kenya	
Buffalo clone M3.7	Kenya	
Buffalo clone M3.9	Kenya	
Buffalo clone M30.2	Kenya	
Buffalo clone M30.5	Kenya	
Buffalo clone M30.8	Kenya	
Buffalo clone M30.11	Kenya	
Buffalo clone M42.2	Kenya	
Buffalo clone M42.5	Kenya	
Buffalo clone M42.8	Kenya	
Buffalo clone M42.12	Kenya	
Buffalo clone 6998.9	Kenya	
Buffalo clone 6998.11	Kenya	
Buffalo-assoc. clone N33.1	Kenya	Buffalo-associated cattle *T. parva* clones
Buffalo-assoc. clone N33.3	Kenya	
Buffalo-assoc. clone N33.4	Kenya	
Buffalo-assoc. clone N33.5	Kenya	
Buffalo-assoc. clone N43.1	Kenya	
Buffalo-assoc. clone N43.3	Kenya	
Buffalo-assoc. clone N43.5	Kenya	
Buffalo-assoc. clone N43.6	Kenya	

*ILRI International Livestock Research Institute, Nairobi, Kenya

When trialled on the test panel, the nested p104 PCR amplified a product of expected size (277 bp) from all of the diverse *T. parva* isolates. However there

was also faint cross-amplification of one *T*. sp. (buffalo) clone (*T*. sp. (buffalo) 6998 clone 10) (Figure 3.1).

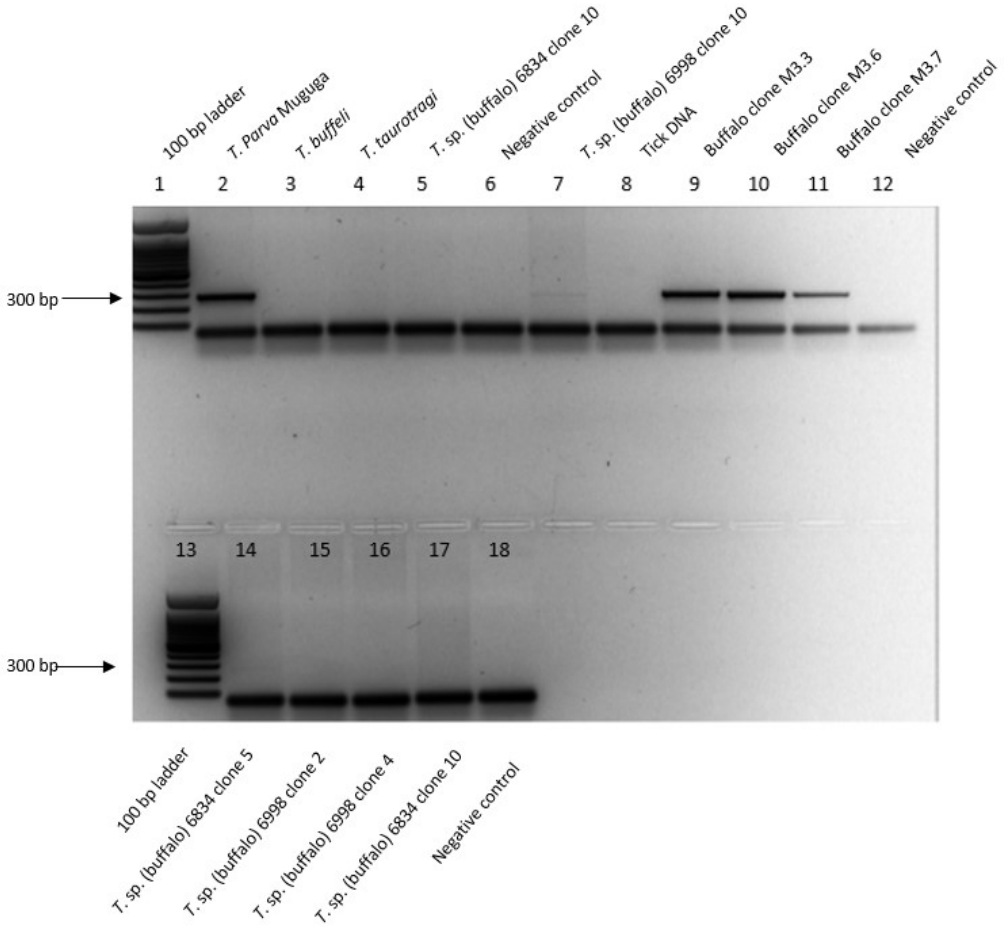

Figure 3.1: Representative agarose gel of PCR amplification using p104 nPCR and DNA from test panel
Weak cross-amplification of *T*. sp. (buffalo) 6998 clone 10 can be visualised in well 7. Strong amplification of *T. parva* Muguga, Buffalo clone M3.3, Buffalo clone M3.6 and Buffalo clone M3.7 can be seen in wells 1, 9, 10 and 11 respectively.
1 = ladder (100 bp), 2 = *T. parva* Muguga, 3 = *T. buffeli*, 4 = *T. taurotragi*, 5 = *T*. sp. (buffalo) 6834 clone 10, 6 = Negative control (no template), 7 = *T*. sp. (buffalo), 8 = Tick DNA, 9 = Buffalo clone M3.3, 10 = Buffalo clone M3.6, 11 = Buffalo clone M3.7, 12 = Negative control (no template), 13 = ladder (100 bp), 14 = *T*. sp. (buffalo) 6834 clone 5, 15 = *T*. sp. (buffalo) 6998 clone 2, 16 = *T.* sp. (buffalo) 6998 clone 4, 17 = *T.* sp. (buffalo) 6834 clone 10, 18 = Negative control (no template)

It was therefore decided to investigate the development of an alternative assay with the aim of improving specificity. Conserved regions of *T. parva* genes

Tp4 (TP03_0210) and Tp5 (TP02_0767) were selected as possible alternative targets. Primers had been designed by Hemmink (2014) that would amplify a 402 bp region of Tp4 and a 327 bp region of Tp5, and so further primers were designed to amplify regions flanking the existing primers in order to develop nested PCR assays for each gene. Details of primer design are described in section 3.3.4. Primer combinations that successfully amplified a product of correct size were taken forward and trialled as a nested assay, using the Hemmink primers as second round 'inner' primers. For each of Tp4 and Tp5 a single nested combination of 'outer' primers successfully amplified their respective gene segment in *T. parva* Muguga, and so these primers were next trialled for sensitivity, using serial dilutions (1:10 – 1:1,000) of *T. parva* Muguga as starting DNA template. Tp4, Tp5 and p104 primers all amplified to 1:1000, however when trialled on the test DNA panel, the Tp5 nPCR showed some cross-amplification with alternative *T.* species, namely *T. buffeli*, *T. taurotragi* and *T*. sp. (buffalo) 6834 clone 5, and so Tp5 was discounted as a potential diagnostic gene. Comparison of Tp4 and p104 was continued with a further serial dilution (1:50 – 1:781,250) of first round template. The p104 primers generated a detectable product from the first round template to a dilution of 1:156,250. Tp4 primers amplified a detectable product from the first round template to a dilution of 1:31,250 (Figure 3.2). Additionally, there was a smaller and second band that amplified less efficiently in the Tp4 nPCR (which may have affected final sensitivity; Figure 3.2), whereas the p104 nPCR resulted in one single band. Based on the superior sensitivity and specificity in terms of a single band of expected size at a high dilution of template, the p104 nPCR assay was selected to take forward in the pipeline, while acknowledging the minor cross-amplification from one *T*. sp. (buffalo) isolate (see section 3.4.2 for further assessment of this aspect).

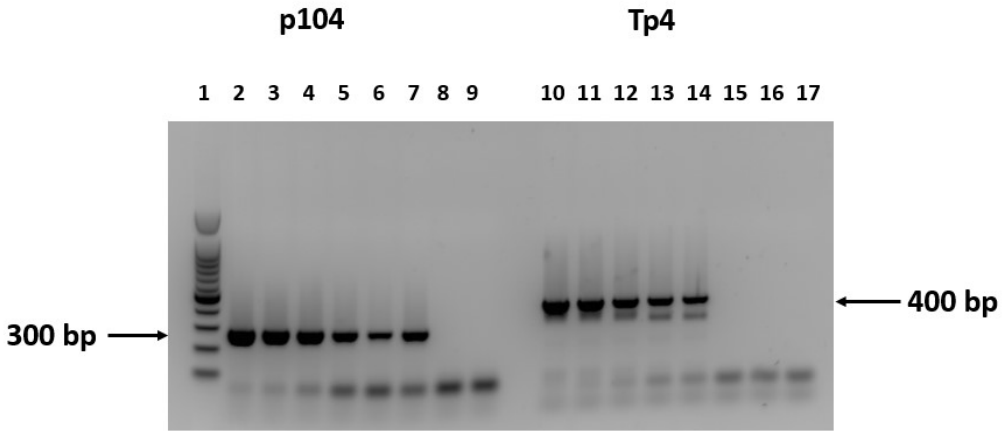

Figure 3.2: Agarose gel amplification of serial dilutions of *T. parva* Muguga comparing p104 and Tp4 primers.
p104 primers were more sensitive, amplifying to a dilution of 1:156,250 compared to Tp4 primers which amplified to a dilution of 1:31,250.

Wells 2-9 are p104 amplicons and wells 10-17 are Tp4 amplicons:
1 = ladder (100 bp), 2 = *T. parva* Muguga 1:50, 3 = 1:250, 4 = 1:1250, 5 = 1:6250, 6 = 1:31,250, 7 = 1:156,250, 8 = 1:781,250, 9 = Negative control (no template)
10 = 1:50, 11 = 1:250, 12 = 1250, 13 = 1:6250, 14 = 1:31,250, 15 = 1:156,250, 16 = 1:781,250, 17 = Negative control (no template).

3.4.2 Primer design

Primers were designed to amplify full-length or near full-length sequences of a number of antigen-encoding genes**.** Primers were required to be specific for *T. parva* as well as being able to amplify all alleles of these genes within the species.

In previous studies by Hemmink (2014) primers were available that amplified the full-length Tp1 gene (Pelle et al. 2011), but this was not the case for Tp2 and Tp9 due to the variability of their sequences and restriction in terms of product length that could be applied to Roche 454 sequencing. Tp2 products could only be produced from all stocks using degenerate primers and there appeared to be variable amplification of different alleles. It was previously not possible to identify primers for Tp9 that would produce the desired size of product appropriate for Roche 454 sequencing from all alleles (Hemmink et al. 2016). However, data became available from the sequenced genomes of

seven *T. parva* isolates (personal communication, Joana Silva), which enabled a renewed attempt to design primer sequences from relatively conserved flanking regions across the selected antigens (Table 3:3), including Tp2 and Tp9. Multiple primers were designed for these flanking regions and subjected to extensive testing for specificity and the ability to detect diverse alleles, using the panel of DNA samples described previously. Existing primers, designed by Pelle for Tp1, Tp2 and Tp9 (2011), MacHugh for Tp9 (unpublished), Silva for Tp2 and Tp9 (unpublished), Hemmink for Tp4, Tp5 and Tp6 (2014) and Nene for p67 (1996) were also tested. A total of twelve target antigen genes were investigated (Table 3:3).

Table 3:3: Characteristics of target genes

Gene	Accession Number	CD8[1]	CD4[2]	Function	Length (bp)
Tp1	TP03_0849	+		Hypothetical protein	1771
Tp2	TP01_0056	+		Hypothetical protein	1019
Tp4	TP03_0210	+		T-complex protein 1 subunit beta	1740
Tp5	TP02_0767	+		Translation initiation factor eIF-1A	527
Tp6	TP01_0188	+		Prohibitin	874
Tp9	TP02_0895	+	+	Hypothetical protein	1005
Tp12	TP01_1091	+		Hypothetical protein	1976
N10/Tp14	TP01_1074		+	Hypothetical protein	1326
N43	TP01_0987		+	Hypothetical protein	3025
A14	TP03_0263	+		Hypothetical protein	2367
N60/Tp16	TP01_0726		+	Elongation factor 1 alpha	1347
p67[3]	TP03_0287			Hypothetical protein	2130

[1]Encode antigens recognised by CD8 T cells from immunised animals
[2]Encode antigens recognised by CD4 T cells from immunised animals
[3]p67 encodes a sporozoite surface antigen recognised by antibodies from immunised animals

Reference genome sequences were accessed from the National Centre for Biotechnology Information (NCBI) Gene Database as well as PiroplasmaDB Genomics Resource (Aurrecoechea et al. 2010) and imported into CLC Main Workbench 7.7 software (QIAGEN). Gene reference sequences were

imported in FASTA file format into Primer3 Input (http://primer3.ut.ee/) (Koressaar et al. 2018, Koressaar and Remm 2007, Untergasser et al. 2012) and primers were designed for both forward and reverse regions. The general restrictions applied were primer length of 18-23 base pairs with a GC content between 30-70%. Primer sequences were then pasted into the BLAST tool in PiroplasmaDB in order to compare to the *T. parva* Muguga reference genome to check for compatibility, and to check that the primers were specific for the particular regions only and did not match elsewhere in the genome. Multiple primer combinations were then ordered (Eurofins Scientific) and reconstituted to their required pmol/µl concentration before trialling on the reference strain, *T. parva* Muguga. Redesign of primers was required for all genes with the exception of Tp1, the most common reason being non-specific amplification of products from other *Theileria* species, *Theileria*. sp. (buffalo) in particular.

It was impossible to identify primers that would amplify Tp2, Tp5, Tp6, Tp9, Tp12, N10, N43 and p67 with adequate sensitivity and/or specificity (Table 3:4).

Table 3:4: Target genes that did not have successful primer design

Gene	Accession number	Reason for failure
Tp2	TP01_0056	Non-specific binding and lack of sensitivity
Tp5	TP02_0767	Non-specific amplification
Tp6	TP01_0188	Non-specific amplification
Tp9	TP02_0895	Inadequate sensitivity (not amplifying the majority of test panel)
Tp12	TP01_1091	Incorrect product size
N10	TP01_1074	Non-specific amplification
N43	TP01_0987	Inadequate sensitivity (not amplifying the majority of test panel)
p67	TP03_0287	Lacking sensitivity and has non-specific amplification

Specific primer combinations were confirmed for Tp1 (Pelle et al. 2011) and were successfully developed for Tp4, A14 and N60.

To further assess sensitivity, primers were tested on field samples that were PCR-positive for *T. parva* (using the diagnostic p104 assay), as well as on a titration series of *T. parva* Muguga-infected cells in uninfected Concanavalin A-stimulated lymphoblast cells. The series was comprised of neat *T. parva* (100%), 0.8% *T. parva*, 0.6% *T. parva*, 0.4% *T. parva*, 0.1% *T. parva* and zero *T. parva* (0%). DNA was extracted from the infection series and concentrations were standardised to 60 ng/µl. In conjunction with this 'infection gradient' 10-fold serial dilutions of *T. parva* Muguga and *T*. sp. (buffalo) were set up. As the test field samples from which DNA was obtained are likely to have contained low percentages of infected cells, the aim of these experiments with a range of known concentrations of infected cells was to obtain information on sensitivity of the PCR assays (on a per-parasite basis) for detecting *T. parva*, including the relative sensitivity in detecting cross-priming with *T*. sp. (buffalo). Tp1 primers detected *T. parva* at 0.8% infection and to a dilution of 1:1000, with no detection of *T*. sp. (buffalo). A14 primers detected *T. parva* at 0.4% infection and to a dilution of 1:1000, and *T*. sp. (buffalo) was detected only when neat. N60 primers detected *T. parva* at 0.1% and to a dilution of 1:1000, with the detection of *T*. sp. (buffalo) at a dilution of 1:10. Thus, it was determined that the primers were more sensitive to the amplification of *T. parva*, detecting *T. parva* at lower concentrations compared to *T*. sp. (buffalo).

When tested on field samples, PCR products were not detected from most samples using single-round PCR, suggesting insufficient primer sensitivity compared to the nested p104 assay. Therefore, nested primers were designed for Tp1 (the primers from Pelle (2011) were also included in the testing process for comparison), Tp4, A14 and N60 in order to increase sensitivity.

Successful amplification of PCR products with the nested assays was confirmed for Tp1 (Figure 3.3), and achieved for Tp4 (Figure 3.4), A14 (Figure 3.5) and N60 (Figure 3.6) (Table 3:5); positive results were obtained for all of the *T. parva* (p104) positive field samples tested. Tp1, A14 and N60 were selected to take forward for trial of the PacBio sequencing pipeline. The

inclusion of Tp1 was considered useful as a ‘control’ to assess the pipeline, as data published by Pelle et al (2011) showed increased diversity of Tp1 in buffalo parasites compared to cattle, providing a reference dataset upon which to compare data generated in this study.

Figure 3.3: Tp1 gene and primer binding sites
Blue represents the gene coding region and green represents gene flanking regions. Red arrows indicate first round primers and grey arrows indicate second round primers (Pelle et al. 2011). Final product size is 1618 bp.

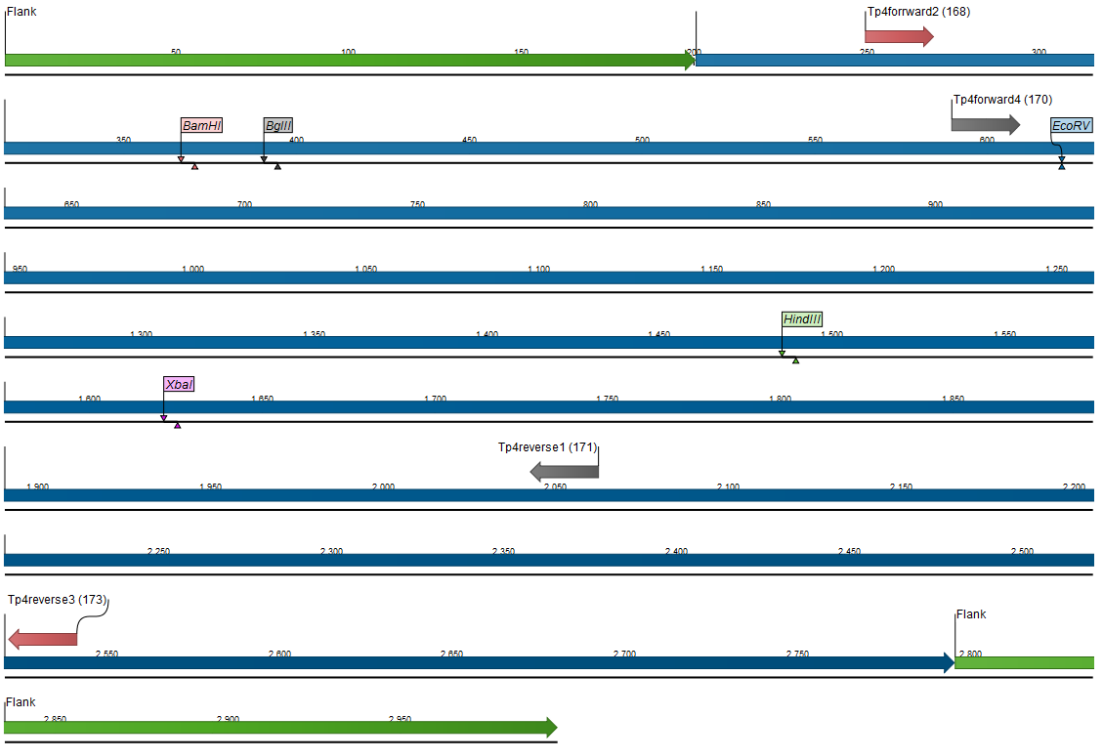

Figure 3.4: Tp4 gene and primer binding sites
Blue represents the gene coding region and green represents gene flanking regions. Red arrows indicate first round primers and grey arrows indicate second round primers. Final product size is 1473 bp.

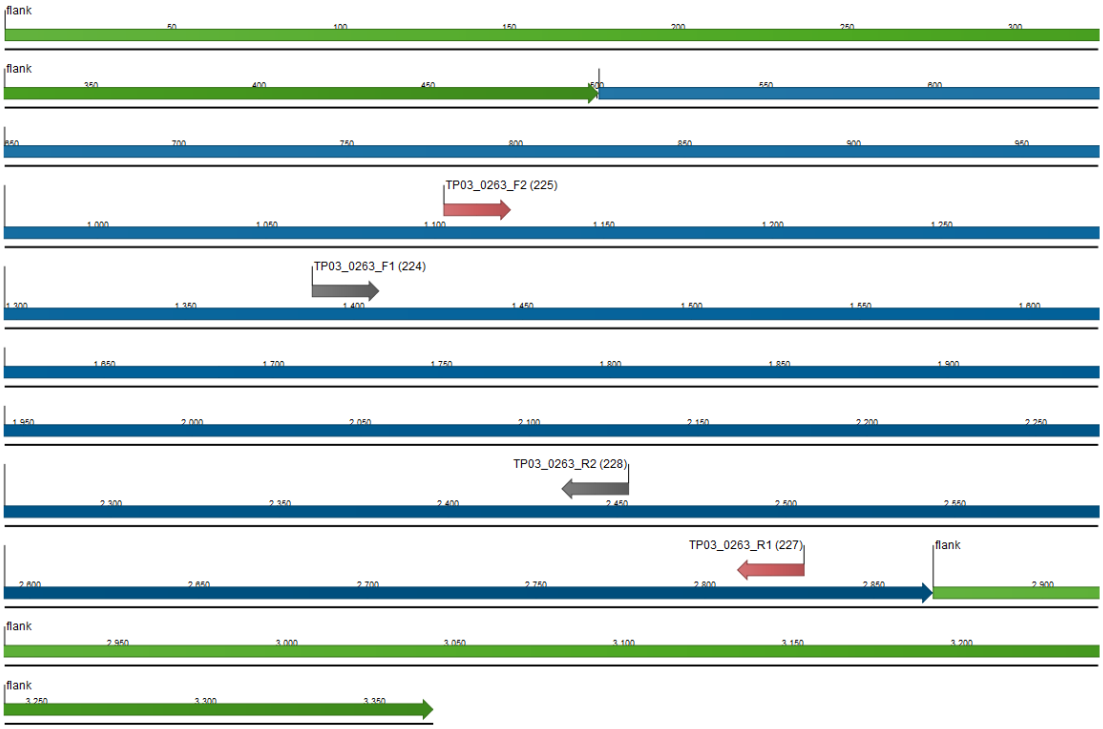

Figure 3.5: A14 gene and primer binding sites
Blue represents the gene coding region and green represents gene flanking regions. Red arrows indicate first round primers and grey arrows indicate second round primers. Final product size is 1066 bp.

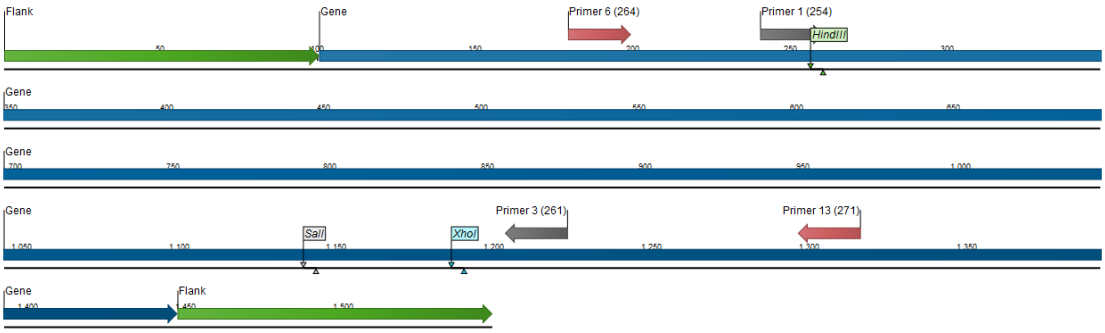

Figure 3.6: N60 gene and primer binding sites
Blue represents the gene coding region and green represents gene flanking regions. Red arrows indicate first round primers and grey arrows indicate second round primers. Final product size is 983 bp.

Table 3:5: Primers and PCR cycling conditions for candidate antigens

Gene	Primers	Cycling conditions
Tp1		
Round 1	F GCTACGCGGAAATCTAGGCT R CATCGTTTGCCAGCACTATGA	98°C for 30 s, 40 cycles (98°C for 10 s, 56°C for 20 s, 72°C for 2.5 min), 72°C for 2 min
Round 2*	F AGGGTCAAAAAAGTTTTATTA R TTAATTTTTGAGGTAAATTTTG	98°C for 30 s, 37 cycles (98°C for 10 s, 54°C for 20 s, 72°C for 1.5 min), 72°C for 2 min
Tp4		
Round 1	F ATACATCCCAAGGCCAAGCT R GGAAGGGGTTGGATAGTGCT	98°C for 30 s, 40 cycles (98°C for 10 s, 58°C for 20 s, 72°C for 2.5 min), 72°C for 2 min
Round 2	F TTACTCATCCTGCCGCTTCT R TGACCTCCACCTCTCAACAC	98°C for 30 s, 30 cycles (98°C for 10 s, 64°C for 20 s, 72°C for 2min), 72°C for 2 min
A14		
Round 1	F ACCAGGCGTTGATGAGATGA R ACTTTGGTTTGTCGCCTGTC	98°C for 30 s, 40 cycles (98°C for 10 s, 56°C for 20 s, 72°C for 2 min), 72°C for 2 min
Round 2	F TACGGGAGCTGTTGAACCTT R GCCTGATGCCGCGTTAATAA	98°C for 30 s, 32 cycles (98°C for 10 s, 60°C for 20 s, 72°C for 1 min), 72°C for 2 min
N60		
Round 1	F TGATCTACAAGCTCGGTGGA R GCGGGTATTCTGTGAAGGTC	98°C for 30 s, 35 cycles (98°C for 10 s, 68°C for 20 s, 72°C for 1.5 min), 72°C for 2 min
Round 2	R AGACATGGGAAAGGGAAGCT F CCTCCAGTGTCTTTCCGGTA	98°C for 30 s, 32 cycles (98°C for 10 s, 56°C for 20 s, 72°C for 1.5 min), 72°C for 2 min

*(Pelle et al. 2011)

3.4.3 Species-specific 18S PCR

During primer validation, there was frequent cross-amplification of *T*. sp. (buffalo). A study by Bishop et al. (2015), based on previous studies by Gubbels et al. (1999) and Oura et al. (2004b), described species-specific primers designed to differentiate *T. parva* and *T*. sp. (buffalo). The RLB-R2 reverse primer (CTAAGAATTTCACCTCTGACAGT) and the species-specific primers (*T. parva*-specific forward primer CTTATTTCGGACGGAGTTCG and *T*. sp. (buffalo)-specific forward primer CGCTTATTTCAGACGGAGTTA) previously described were used in an alternative step down PCR assay (unpublished - personal communication Hanneke Hemmink, ILRI, Kenya).

Cycling conditions were as follows: 94°C for three minutes, five cycles step-down (94°C for 30 seconds, 58-52°C for 30 seconds, 72°C for 30 seconds), 35 cycles of (94°C for 30 seconds, 52°C for 30 seconds, 72°C for 30 seconds), 72°C for 10 minutes. Step down PCR uses progressive cycles of incremental reduction in annealing temperature and is used to increase specificity (Hecker and Roux 1996). This assay was able to differentiate *T. parva* from *T.* sp. (buffalo) confirming the observed cross-amplification of the Tp2 primers with *T*. sp. (buffalo) 6998 clone 10 DNA to be genuine and not a result of, for example, contamination with *T. parva* DNA. It was decided to sequence the cross-amplifying amplicons to assess whether the *T*. sp. (buffalo) alleles were sufficiently distinct from *T. parva* sequences to be differentiated bioinformatically in the event that a field sample contained both species.

3.4.4 Sanger sequencing of PCR products from *T*. sp. (buffalo)

For Tp2, primers designed by the laboratory of Joana Silva (University of Maryland, unpublished) were tested and amplified the majority of the test panel (21 of 23 *T. parva* isolates), but these primers also amplified two *T*. sp. (buffalo) stocks [*T*. sp. (buffalo) 6834.10 and *T*. sp. (buffalo) 6998.10]. As already discussed in 3.4.3 this cross-reaction was confirmed to be genuine by species-differentiating PCR. It was thought that if *T. parva* Muguga and *T*. sp. (buffalo) could be differentiated bioinformatically then it may be feasible to continue use of the Silva primers for amplification of Tp2. Plasmid DNA amplicons were sequenced at DNA Sequencing and Services, University of Dundee, using standard primers M13 Rev and T7. Sequence data showed extensive homology with only a few mismatches between *T. parva* and *T*. sp. (buffalo); without sequencing alleles from multiple *T. parva* and *T*. sp. (buffalo) isolates to establish consistent mismatches, differentiation of species was impossible. It was therefore decided to exclude the Silva Tp2 primers.

3.5 Samples for long-read sequencing

DNA samples were from cattle and buffalo in the Ol Pejeta game conservancy, Laikipia District, Kenya. Cattle (n=30) and buffalo DNA samples (n=12) were kindly provided by Dr Philip Toye and Dr Annie Cook (ILRI, Kenya).

Samples from this site were used based on findings from previous studies in Ol Pejeta, by Sitt et al (2015) and Hemmink et al (2018), which collected blood samples from eight buffalo; the *T. parva* parasites in these samples represented persistent infections in buffalo. In the Sitt study a group of twelve naïve cattle were introduced into an area on the conservancy that had not been grazed by cattle for several years and thus infections in the cattle would be from ticks that had fed on buffalo, i.e. were buffalo-derived. All of the cattle developed infection and clinical signs of disease and most died or were euthanased. Blood samples from these cattle represent acute buffalo-derived infections. In this project, using samples from the defined setting of this study site allows the diversity of *T. parva* in buffalo to be compared with that of parasites acquired by sentinel cattle introduced into the area where the buffalo were grazed and so makes for an ideal sample set from a livestock/wildlife interface on which to work up the sequencing pipeline for application to the larger set of field samples from the principal study area of the project, the boundary of the Serengeti National Park, Tanzania.

3.5.1 Sample preparation for sequencing

The diagnostic nPCR assay described was applied to all Ol Pejeta buffalo and cattle DNA samples to confirm *T. parva* infection status. Of the p104 positive samples, two buffalo and two cattle were selected to be used in a trial of the PacBio sequencing and bioinformatics pipeline; as this work was an exploratory study to test the suitability of the sequencing technology for detecting diversity, a minimal sample set was sufficient to test the utility of PacBio, as well as working up the analytical pipeline. Amplification of all three target antigen genes – Tp1, A14 and N60 – in both selected buffalo (D4 and

D7) and cattle samples (2684 and 2732) was carried out using nested PCR (as detailed in Section 3.4.2), creating a total of 12 DNA amplicons that were prepared for PacBio sequencing. PCR purification of amplicons was carried out, as previously described.

In order to ensure high quality starting material and to meet the requirements for library preparation, a series of quality checks was carried out on all amplicons (Table 3:6). DNA purity was assessed by NanoDrop spectrophotometry (2016 Thermo Fisher Scientific) where a 260/280 ratio of ~1.8 and a 260/230 ratio of 2.0-2.2 are considered thresholds of sufficient purity for downstream library preparation. Qubit 3.0 fluorimetry (Invitrogen) was used to measure DNA concentration. Agilent 2200 TapeStation System (Agilent Technologies) was used to assess DNA integrity and size, using Agilent D5000 Screen Tape System. 200-500 ng per amplicon sample was required for library preparation.

Table 3:6: Amplicon sample information

Amplicon	Species & antigen	Concentration (ng/µl)	260/280 ratio	260/230 ratio	Size (bp)
C2	Cow Tp1	13.5	1.9	1.8	1618
C5	Cow Tp1	19.6	1.6	1.3	1618
B2	Buffalo Tp1	22.8	1.8	1.6	1618
B4	Buffalo Tp1	11.7	2.4	2.9	1618
C7	Cow A14	52.0	1.9	2.3	1066
C10	Cow A14	51.8	1.9	2.2	1066
B7	Buffalo A14	13.8	1.7	1.6	1066
B9	Buffalo A14	74.8	1.8	2.1	1066
C12	Cow N60	21.8	1.7	2.4	983
C15	Cow N60	84.1	1.9	2.0	983
B12	Buffalo N60	23.8	1.7	1.9	983
B14	Buffalo N60	79.4	1.9	2.1	983

Amplicons were submitted to the Centre for Genomic Research, University of Liverpool, where they were multiplexed and sequenced on a single SMRT cell on a PacBio RS II machine.

3.6 PacBio sequencing technology

In order to obtain full-length or near full-length sequences, single-molecule real-time (SMRT) sequencing, third-generation sequencing (TGS) technology developed by PacBio (Rhoads and Au 2015), was utilised, which sequences single DNA molecules (Schadt et al. 2010). Single-molecule sequencing (SMS) has increased the rate of sequencing and increased read lengths. However, it has a lower throughput and a higher rate of error of ~11% (Korlach 2015) when compared to second-generation sequencing (SGS) (Rhoads and Au 2015).

PacBio technology circularises the template DNA by the addition of hairpin adaptors at each end, creating a SMRTbell (Figure 3.7). As the SMRTbell is a closed circle the polymerase can replicate a single strand of the double-stranded DNA (dsDNA) target, and then continue to incorporate bases of the hairpin adaptor and the other strand, and will continue to do so continuously for the lifespan of the polymerase, thereby sequencing the target multiple times and providing multiple coverage of a single sequence. The sequencing of a single strand of a SMRTbell, without any adaptor sequence is termed a “subread”. When a subread starts and finishes at an adaptor it is termed a “full-pass subread”. Multiple subreads aligned create a circular consensus sequence (CCS) (Rhoads and Au 2015), and provide the ability (with sufficient subreads) to correct for the initial random PacBio error and gain accurate sequence information.

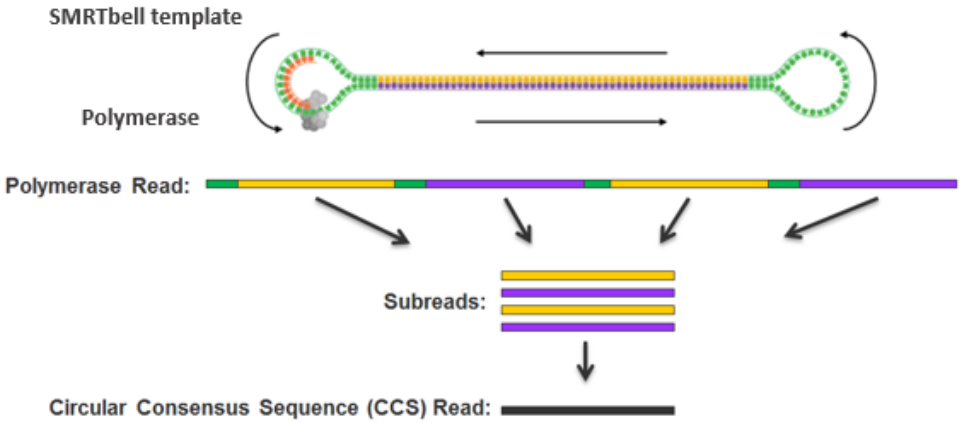

Figure 3.7: Annotation of SMRTbell template
SMRTbell (yellow and purple) with hairpin adaptors (green) creating a closed circle. The DNA polymerase (grey) is fixed in a ZMW nanowell and bases are incorporated into the read strand (orange). Multiple subreads (sequences from a single pass of polymerase with no adaptor sequence) are generated, and then aligned to create the circular consensus (CCS) read (black). Image adapted with permission from Pacific Biosciences

SMRTbell template is applied to a SMRT cell, a chip containing sequencing units that allow light to be observed in a very small volume; these units are called zero-mode waveguides (ZMW) and there are 150,000 in each SMRT cell of a RS II sequencer. At the base of each ZMW there is a single polymerase attached, which binds to a hairpin adaptor to start replication. Fluorescently-labelled nucleotides, each emitting distinct spectrums, are added to the SMRT cell. Light pulses are produced, identifying each base. A dye-linker-pyrophosphate product cleaves from its nucleotide, diffusing out of the ZMW and terminating the fluorescence (Eid et al. 2009). The fluorescent pulses in all of the ZMWs are recorded as a real-time "movie" and interpreted as a continuous long read (CLR) sequence of bases (Rhoads and Au 2015). Chapter 4 utilises the Sequel sequencer, which has increased throughput, with one million ZMWs, able to produce approximately 500,000 reads per SMRT cell, as well as increased read length and accuracy (PacBio 2015-2019).

3.6.1 Bioinformatics methodology

Submitted amplicons were individually barcoded, pooled and multiplexed for sequencing by the Centre for Genomic Research (CGR). Raw sequencing data were received in HDF format – these files contained basecalling information as well as metadata pertaining to the sequencing run (PacBio 2015-2019).

Data were demultiplexed and CCS generated using the SMRT Analysis Portal (v2.3.0). Data were converted into FASTQ format and downloaded. FASTQ format contains quality score information which allows for generation of quality metrics. Siddharth Jayaraman (bioinformatician in L. Morrison group) provided the FASTQ files with quality control (QC) metrics. Read length distribution was checked, ensuring each amplicon fell into the correct bp range. The coverage was assessed by the number of full passes. The processing and filtering steps described were carried out by Siddharth Jayaraman;

Filtering was carried out in a step-wise manner:

1. Size selection and strand-correction. All the reads were required to be in the same orientation for clustering.
2. Primer detection. The presence of a forward and reverse primer at each end was required, ensuring full-length reads.
3. Full passes. A minimum of ten full passes of each sequence was selected to allow confidence in read quality by allowing removal of PacBio random errors introduced at the initial sequencing stage.

Multiple tools were used for the filtering steps:

PRINSEQ software (Schmieder and Edwards 2011) was used to read filter, allowing manipulation of FASTA files to see the read length distribution and removing sequences that were not full length.

A R script (Jayaraman 2018a) was used for strand-correction and a Matlab script (Jayaraman 2018b) was used for primer detection and CCS generation.

BLAST was used to check amplicon expected length and strand-orientation. A local blast of sequences was run against the *T. parva* Muguga genome reference sequence for each antigen.

Reads were then sorted by length to aid clustering, using SeqKit (Shen et al. 2016). CD-HIT programme (Huang et al. 2010) was used to address the aim of assessing antigenic diversity, grouping reads by percentage similarity. CD-HIT used the processed reads to cluster sequences at a percentage similarity ranging from 80-99% similarity.

3.6.2 Bioinformatics results

A total of 38,730 reads were received. Due to suboptimal PacBio RS II loading efficiency (Eid et al. 2009) between 35,000 and 70,000 (23-46%) of the 150,000 ZMWs on each SMRT cell typically provide reads (Rhoads and Au 2015).

Read lengths of pooled inserts showed a peak of expected size for each antigen gene (Figure 3.8).

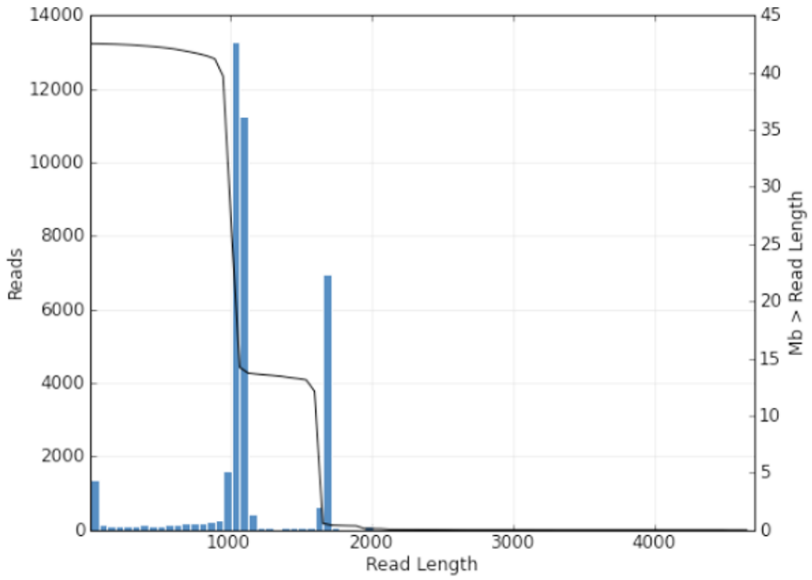

Figure 3.8: Read lengths of inserts
Expected amplicon sizes were 983 bp (N60), 1,066 bp (A14) and 1,618 bp (Tp1).

There was a mean of 26 full passes (Figure 3.9). Although PacBio sequencing can have a relatively high error rate, of approximately 11% for single pass reads (Korlach 2015), the errors are random and so when consensus is created with multiple passes the error rate is reduced; 15 passes produces a median accuracy of 99.3% (Eid et al. 2009).

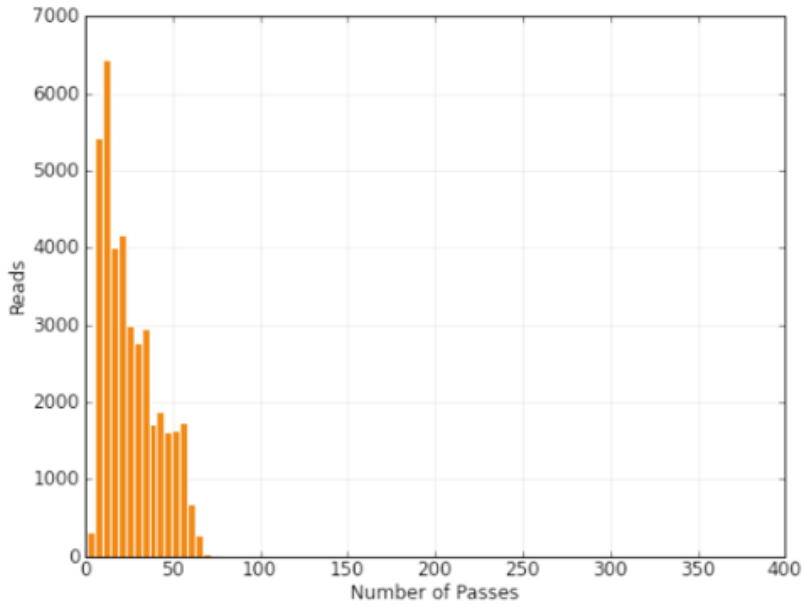

Figure 3.9: Distribution of number of full passes per read (mean = 26)

For this analysis 10 full passes was used as a filter, providing an accuracy of greater than 99% (Korlach 2015).

Reads were lost in the filtering process, but by using a stringent filtering approach those that remain represent high confidence data (Table 3:7).

Table 3:7: Read number per antigen and sample at each data filtering step

	Sample	Raw reads	Size-selected & Strand-corrected	Primer-detected	Final processed reads
Tp1	Cow 1	1,936	1,852	1,076	1,076
	Cow 2	1,747	1,612	852	852
	Buffalo 1	2,608	2,436	1,234	1,234
	Buffalo 2	2,488	2,329	1,274	1,274
A14	Cow 1	3,518	3,417	1,556	1,556
	Cow 2	4,457	4,210	2,157	2,157
	Buffalo 1	1,539	1,512	712	712
	Buffalo 2	3,001	2,909	1,586	1,586
N60	Cow 1	2,480	2,423	1,521	1,521
	Cow 2	5,131	5,030	3,401	3,401
	Buffalo 1	2,671	2,618	1,843	1,843
	Buffalo 2	5,253	5,108	3,184	3,184

Processed reads were clustered for similarity (Table 3:8). Clusters were formed on the basis of percentage similarity and a count of the groups formed a proxy for diversity, indicating the level in each animal.

Table 3:8: Numbers of clusters for each antigen and sample
identified using CD-HIT ranging from 80-99% similarity thresholds

		CD-HIT cluster counts (% similarity)											
Antigen	**Sample**	**80**	**85**	**90**	**91**	**92**	**93**	**94**	**95**	**96**	**97**	**98**	**99**
Tp1	Cow 1	4	4	4	5	5	5	5	5	5	8	14	55
	Cow 2	3	3	4	4	4	4	4	4	4	6	16	83
	Buffalo 1	2	3	3	3	3	3	3	3	3	5	13	114
	Buffalo 2	4	4	4	4	5	5	5	6	6	9	22	131
A14	Cow 1	1	1	2	2	2	2	2	2	2	2	3	13
	Cow 2	1	1	2	2	3	3	3	3	3	3	5	24
	Buffalo 1	1	1	1	1	1	1	1	1	1	1	1	5
	Buffalo 2	1	2	2	2	2	2	2	2	2	2	2	8
N60	Cow 1	1	2	2	2	2	2	2	2	2	3	3	22
	Cow 2	1	1	1	1	1	1	1	1	1	1	3	32
	Buffalo 1	1	1	1	1	1	1	1	1	1	1	2	17
	Buffalo 2	1	1	1	1	1	1	1	1	2	2	5	33

The clustering data indicate that for all antigens a threshold of greater than 96% similarity is required to start to resolve allelic differences for Tp1, and above 98% for A14 and N60. It was decided that 99% would be selected for the bioinformatic pipeline, as this enables differentiation of allelic differences across all three antigens. Given the amplicon lengths (983 bp for N60, 1,066 bp for A14 and 1,618 bp for Tp1), this threshold represents approximately 10 (N60), 11 (A14) and 16 (Tp1) single nucleotide polymorphisms between sequences required to define a distinct allele. This is consistent with the degree of polymorphism previously identified between, for example, allele sequences of Tp1 (Pelle et al. 2011). The degree of polymorphism clearly differed between antigens, with Tp1 being significantly more polymorphic at the 99% threshold (between 55 and 131 alleles per animal) than either A14 (between 5 and 24 alleles per animal) or N60 (between 8 and 33 alleles per animal).

Singleton reads, defined as clusters at the 99% similarity threshold represented by a single read, were removed, as the presence of two independent reads was considered as a requirement to improve confidence in

the identity of a particular cluster. Combined with the initial data filtering steps, the resultant identified clusters should provide robust, but probably relatively conservative, estimations of allelic diversity.

Tp1 clusters at 99% similarity

A total of 4,436 reads were obtained for Tp1 across the four samples (Table 3:9), which resulted in 282 clusters at the 99% similarity threshold. Removal of singleton reads reduced the total number of reads by only 258 (5.8%), but reduced the cluster number to 109, indicating 61.3% of clusters were represented by a single read (at the data thresholds used).

Table 3:9: Tp1 processed reads with and without singleton reads

	Total	Cow 1	Cow 2	Buffalo 1	Buffalo 2
Processed reads (including singletons)	4,436	1,076	852	1,234	1,274
Clusters	282				
Processed reads (excluding singletons)	4,178	1,038	796	1,159	1,185
Clusters	109	29	44	61	69

Of the total 4,178 processed reads, with singletons removed, at the 99% similarity threshold, 4,080 were common to cattle and buffalo, 44 were unique to cattle and 158 were unique to buffalo. Removal of singleton reads resulted in 109 clusters (Figure 3.10). In cow 1 there was one unique cluster and in cow 2 there were 11 unique clusters. There were 19 clusters unique to buffalo 1 and 26 clusters unique to buffalo 2.

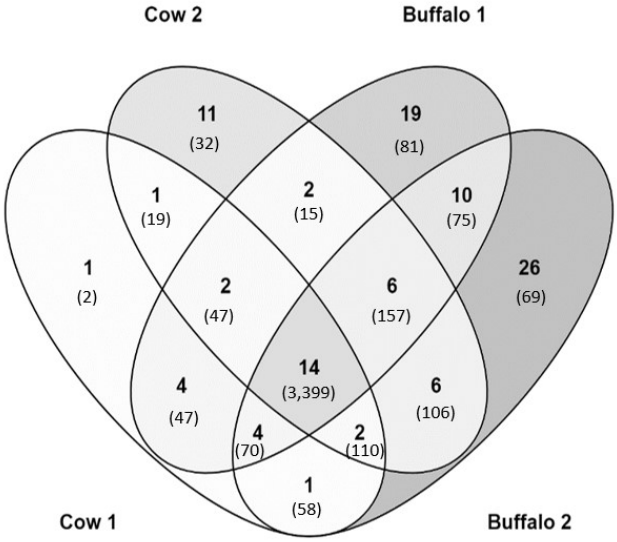

Figure 3.10: Venn diagram of Tp1 clusters, with singletons removed
A total of 109 clusters were present after singleton reads were removed. Read numbers per cluster are shown in brackets. Darker shading indicates higher number of clusters.

A14 clusters at 99% similarity

A total of 6,011 reads as obtained for A14 (Table 3:10), which resulted in 34 clusters at the 99% similarity threshold across the four samples. When singleton reads are removed, the total number of reads was only reduced by 23 (0.3%), and the cluster number was reduced to 23, indicating 32.3% of clusters were represented by a single read (at the data thresholds used). A14 is therefore clearly much less diverse than Tp1, but most allelic diversity is captured by multiple reads.

Table 3:10: A14 processed reads with and without singleton reads

	Total	Cow 1	Cow 2	Buffalo 1	Buffalo 2
Processed reads (including singletons)	6,011	1,556	2,157	712	1,586
Clusters	34				
Processed reads (excluding singletons)	5,988	1,551	2,150	706	1,581
Clusters	23	19	21	10	12

Of the total 5,988 processed reads obtained for A14, with singletons removed, 5,982 were shared between cattle and buffalo, 19 were unique to cattle and zero were unique to buffalo. With singleton reads removed, a total of 23 clusters were identified (Figure 3.11). The sequences from cows 1 and 2 included two and three unique clusters respectively, whereas no clusters were found in the two buffalo.

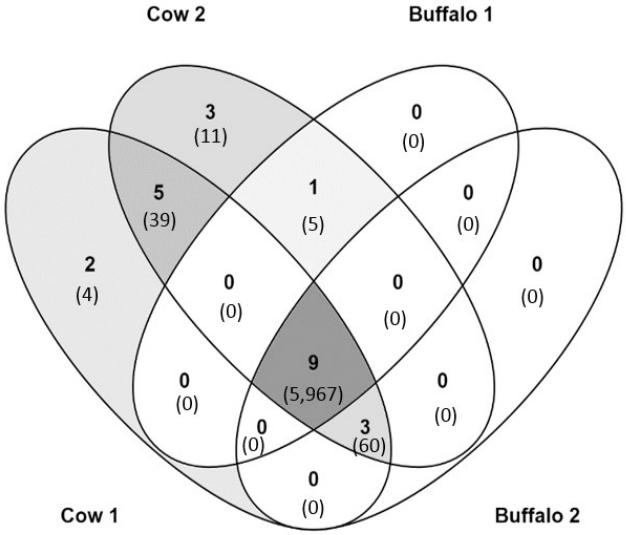

Figure 3.11: Venn diagram of A14 clusters, with singletons removed
A total of 23 clusters were present after singleton reads were removed. Read numbers per cluster are shown in brackets. Darker shading indicates higher number of clusters.

N60 clusters at 99% similarity

A total of 9,949 reads was obtained for N60 (Table 3:11), which resulted in 64 clusters at the 99% similarity threshold across the four samples. Removal of singleton reads reduced the total number of reads by 45 (0.4%), and reduced the cluster number to 37, indicating 42.2% of clusters were represented by a single read (at the data thresholds used). N60 lies between Tp1 and A14 in terms of level of diversity and clusters being represented by single reads.

Table 3:11: N60 processed reads

	Total	Cow 1	Cow 2	Buffalo 1	Buffalo 2
Processed reads (inc singletons)	9,949	1,521	3,401	1,843	3,184
Clusters	64				
Processed reads (excl singletons)	9,926	1,512	3,384	1,834	3,174
Clusters	37	28	32	26	35

Of the total 9,926 processed reads, with singletons removed, 9,914 were shared by cattle and buffalo, 6 were unique to cattle and 6 were unique to buffalo. With singleton reads removed (Figure 3.12), 37 clusters were identified. Cow 1 had zero unique clusters and cow 2 had two unique clusters. Sequences from buffalo 1 had no unique clusters whereas reads from buffalo 2 included four unique clusters.

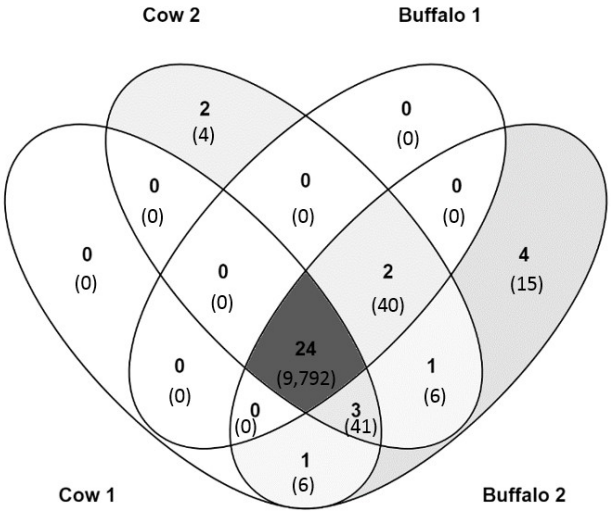

Figure 3.12: Venn diagram of N60 clusters, with singletons removed
A total of 37 clusters were present after singleton reads were removed. Read numbers per cluster are shown in brackets. Darker shading indicates higher number of clusters.

To summarise the results of the bioinformatics pipeline, Tp1 showed more clusters for the buffalo samples than the cattle, with 19 clusters for buffalo 1 and 26 clusters for buffalo 2, compared to one cluster for cow 1 and 11 clusters for cow 2. Overall though the majority of reads were shared between cattle and buffalo, with 4,080 of 4,178 shared. Of the 109 clusters, 41 clusters were shared across cattle and buffalo. 13 clusters were shared between cattle and 55 were shared between buffalo.

A14 showed very few clusters, with 23 in total and more in the cattle samples than the buffalo; cow 1 had two clusters and cow 2 had three clusters, compared to both buffalo with zero clusters. Almost all of the reads were shared between cattle and buffalo, with 5,982 of 5,988 shared. Of the 23 cluster, 13 were shared across cattle and buffalo. 10 clusters were shared between cattle and there were no buffalo clusters.

N60 showed slightly more clusters for buffalo than cattle; cow 1 had zero clusters and cow 2 had two clusters compared to buffalo 1 with zero clusters and buffalo 2 with four clusters. Of the 9,926 reads, 9,914 were shared between cattle and buffalo. Of the 37 clusters, 31 were shared across cattle

and buffalo. 2 clusters were shared between cattle and 4 were shared between the buffalo.

3.7 Discussion

In order to analyse the genetic diversity in *Theileria parva* antigen-encoding genes, a genotyping pipeline was established using PacBio long-read sequencing. A total of twelve antigen-encoding genes were investigated. Of the twelve, seven of the gene products were originally identified through being recognised by CD8 T cells, three gene products through being recognised by CD4 T cells, one was recognised by both CD8 and CD4 T cells and one was known to be a target for neutralising antibodies. These antigens, therefore, are potentially involved in immunological control of *T. parva* by the host. Blood samples were obtained from resident buffalo and from introduced naïve cattle in an area of the Ol Pejeta game conservancy, Kenya, not previously grazed by cattle. Hence, infections in the cattle were expected to be initiated by parasites transmitted from buffalo. PCR amplicons were generated for three antigen-encoding genes from blood samples collected from the animals and PacBio sequencing applied to obtain long-read sequences that were analysed with a purpose-built bioinformatics workflow.

Development of the pipeline required extensive primer design and validation, with each primer set being trialled on a test DNA panel representing intra and inter-species diversity in *Theileria*. This extensive testing was essential to ensure that for the primers that were ultimately intended to be used to amplify *T. parva* from field samples (in which there could be multiple species of *Theileria*, including the very closely related *T*. sp. (buffalo)), primer sensitivity was sufficient so that all allelic variants of *T. parva* were being amplified as well as establishing that the primers were indeed only specific in the amplification of *T. parva.* Many primer sets were tested and discarded due to lack of specificity and/ or sensitivity.

In testing primer specificity, the closely related *Theileria.* sp. (buffalo) proved hugely problematic, frequently cross-amplifying and thus ruling out primer sets. *T*. sp. (buffalo) and *T. parva* are known to be very closely related (Bishop 2015, Chaisi et al. 2011), but *T*. sp. (buffalo) is considered a distinct species (Allsopp et al. 1993, Bishop 2015, Conrad et al. 1987). The species making up the *T.*

taurotragi clade are genetically more closely related than species from other clades (Pienaar et al. 2014). Mans et al. (2011) showed 3-5 nucleotide differences in the 18S region of the genotypes of *T*. sp. (buffalo) and *T. parva*. Hemmink (2014) found only three nucleotide differences between *T. parva* and *T*. sp. (buffalo) sequences over a 375 bp region of 18S rRNA in DNA samples from buffalo in the Kruger National Park, South Africa, and the Ol Pejeta game conservancy, Kenya. Phylogenetic studies of 18S sequences showed distinct clusters of *T*. sp. (buffalo) from *T. parva* sequences (Hemmink 2014) corroborating the likelihood that *T.* sp. (buffalo) is a distinct species.

Little work has been done previously on *T. parva* diagnostics, but new *Theileria* species and genotypes have been discovered in the past ten years and there is increased awareness of the extent of mixed infections in domestic animals (Criado-Fornelio et al. 2004, Mans et al. 2011, Sivakumar et al. 2014). The picture is further complicated by mixed *Theileria* infections in both mammalian hosts as well as tick hosts in areas where wildlife interact with livestock as well as tick vectors (Kariuki et al. 2012, Lawrence et al. 1983). Extensive work-up has been carried out in the work of this chapter to ensure that the p104 assay is a suitably sensitive and specific assay for diagnosing *T. parva* in field samples, specifically where cattle can interact with wildlife and where mixed infections are likely to be common, including with *T*. sp. (buffalo). The minor cross-amplification seen with the p104 nPCR was considered acceptable given the creation of specific primers for the target antigen amplification downstream and the likelihood that such weak cross-reaction is unlikely to be detectable in field samples that contain much lower proportions of infected cells compared to the test experimental samples. The diverse DNA panel was a unique and invaluable resource for validating primer specificity. For future development of diagnostics it is imperative that multiple species and isolates of *T*. sp. (buffalo) are included.

Antigen genes Tp1, Tp2 and Tp9 were initially targets of interest due to their known highly polymorphic nature (Connelley et al. 2011, Graham et al. 2008, Hemmink et al. 2016, Pelle et al. 2011). However, despite extensive primer

design and optimisation, it was not possible to identify primers that would specifically amplify products for either Tp2 or Tp9 – this was due to the extreme polymorphism throughout the sequences of these genes in *T. parva*, such that there were very few conserved regions to design primers, and none that enabled species-specific amplification. Therefore, investigation of these genes had to be abandoned. Successful amplification of Tp1, A14 and N60 was achieved and PacBio long-read sequencing was applied to these amplicons.

Analysis of the data generated in the trial of the pipeline showed that overall the majority of reads were shared between buffalo and cattle species, for all three antigens, suggesting that most, if not all buffalo-maintained *T. parva* genotypes are able to establish acute infections in cattle following tick transmission. Sequence data for the three parasite genes demonstrated varied levels of diversity in the buffalo and cattle samples. Tp1 had multiple variants at a similarity threshold of 99% (109) and showed more unique clusters in buffalo than in cattle. Pelle et al. (2011) demonstrated multiple allelic variants of Tp1, with the majority in buffalo isolates or buffalo-derived isolates, and far fewer variants observed in cattle-derived isolates. Hemmink et al. (2018) also found extensive Tp1 diversity in buffalo in this location.

A key difference in the samples examined in this study, which could potentially affect the level of diversity, is that the parasites in the buffalo were from carrier infections acquired over a prolonged period of time, whereas the cattle samples were from acute infections acquired within the previous 4-5 weeks. In this regard, since buffalo-derived parasites are known to differentiate poorly to the piroplasm stage in cattle and therefore are not readily transmitted by ticks (Schreuder et al. 1977), most of the parasite DNA will be from the schizont stage in cattle whereas most of the parasite DNA from buffalo will be from piroplasms. Nevertheless, in addition to producing few if any piroplasms in cattle, buffalo-derived *T. parva* typically produce fewer schizonts compared to cattle-maintained parasites, and yet results in acute disease (Young et al. 1978).

Antigen gene A14 showed very little diversity, with 23 variants across all four animals, and interestingly more unique clusters in cattle than in buffalo. N60 showed some diversity, with 37 variants across the four animals, and slightly more diversity in buffalo than cattle. Diversity analyses have not been carried out on A14 and N60 previously so the level of diversity expected was unknown.

To summarise, a pipeline was successfully generated to allow the analysis of antigenic diversity in *Theileria parva* field populations, utilising PacBio long-read sequencing technology. Amplicons were produced for three antigen genes – Tp1, A14 and N60 – from *T. parva* positive cattle and buffalo. A multiplexing approach was used in order to maximise sample throughput per SMRT cell. A bioinformatic pipeline was created, using stringent filtering to eliminate, as far as possible, PacBio error and to give confidence to the sequence data. By setting a high CCS filter the PacBio error rate is mitigated by 'washing out' random errors as consensus is created. Generation of this pipeline allows for a more extensive analysis of antigenic diversity in the principal sample set from Serengeti District, Tanzania, that is described in Chapter 4.

Chapter 4: Analysis of antigenic diversity in *T. parva* field populations using long read sequencing

4.1 Introduction

Protozoan parasites are genetically complex and undergo sexual recombination; these features enable considerable genetic and antigenic diversity which can facilitate parasite adaptation within the mammalian host (Morrison 1996). As previously described, when cattle are vaccinated with the Muguga cocktail, there is usually only partial protection against buffalo-derived *T. parva* (Radley 1981, Radley et al. 1979), which is indicative of greater or different antigenic diversity within the buffalo-maintained *T. parva* population (Morrison et al. 2015). Field studies have also shown that vaccinated cattle introduced into areas grazed only by buffalo were fully susceptible to challenge by buffalo-derived parasites (Pelle et al. 2011, Sitt et al. 2015). Due to the relatively recent introduction of indigenous cattle into East and southern Africa, it is reasonable to conclude that cattle acquired *T. parva* parasites from buffalo and that parasite selection in cattle was based on the ability of particular genotypes to differentiate into the tick-transmissible piroplasm stage in the cow host. This then suggests that the limited diversity in cattle *T. parva* is due to genetic constraints in parasite adaptation to tick transmission (Morrison et al. 2015, Oura et al. 2005, Oura et al. 2004a, Oura et al. 2011b).

It is known that components of the Muguga cocktail vaccine can be transmitted from a vaccinated cow to ticks and then onwards to other cattle in the field (De Deken et al. 2007, Geysen et al. 1999). Movements of asymptomatic carrier cattle are an additional source of introducing novel parasite genotypes to cattle populations (De Deken et al. 2007, Hayashida et al. 2013).

Studies on *T. parva* field populations in Uganda have observed the frequent occurrence of multiple alleles in adult cattle samples, demonstrating the presence of mixed *T. parva* genotypes and indicating that the tick vector could be host to multiple *T. parva* genotypes and thus allow for genetic exchange (Oura et al. 2005). Hemmink et al. (2016) also observed many allelic variants within individual buffalo in Kenya and South Africa, using Roche 454

sequencing of the 18S rRNA gene, also indicating mixed genotypes being acquired per infection, or multiple infection 'events'. Oura et al (2011b) examined *T. parva* populations in cattle and buffalo at Lake Mburo National Park in Uganda and observed distinct parasite populations in cattle and buffalo indicating no transmission of buffalo-derived parasites to cattle.

During these studies of parasite population structure, Oura et al. (2005) and Patel et al. (2011) established no particular relationship between parasite genetic similarity and geographical origin, using satellite markers, with distinct parasite genotypes even originating from the same farm. Hemmink et al. (2016) also identified limited geographic differentiation, with significant sequence diversity shared between different geographical populations in Kenya and South Africa.

Several more targeted studies have been carried out to examine the molecular diversity in *T. parva,* using an array of techniques. Monoclonal antibody studies were initially used (Conrad et al. 1987, Minami et al. 1983); Conrad et al. (1989) and Toye et al. (1995) carried out *in vitro* studies on a polymorphic immunodominant molecule (PIM) expressed by *T. parva* and demonstrated that monoclonal antibodies specific for the PIM antigen could inhibit the invasion of lymphocytes by sporozoites. Using the monoclonal antibodies, a cDNA clone encoding the PIM antigen of *T. parva* Muguga was generated and PIM cDNA was also isolated from buffalo-derived *T. parva*. The two cDNAs showed homology in the flanking regions, but sequence variation was observed in the central regions.

Further characterisation studies used gene probes binding to multi-copy genes to detect restriction fragment length polymorphisms (RFLPs) in genomic DNA (Bishop 2002) and the establishment of the *T. parva* genome sequence (Gardner et al. 2005) allowed for DNA satellite markers to be identified, thus allowing for further genotyping studies (Katzer et al. 2011, Oura et al. 2003) and the demonstration of high levels of genomic polymorphism in *T. parva* isolates from multiple geographical locations.

Studies by Graham et al. (2006, 2008) identified *T. parva* antigens recognised by CD8 T cells from immunised animals and some of these, Tp1 and Tp2 in particular, have demonstrated extensive sequence polymorphism and greater diversity in the buffalo-derived *T. parva* population than in the cattle-maintained population (Hemmink et al. 2018, Kerario et al. 2019, Pelle et al. 2011). By studying *T. parva* antigens, it has been possible to investigate their involvement with the immune response and to investigate the selection pressures associated with immunity and antigenic diversity (Pelle et al. 2011).

Studies by Hemmink et al. (2014, 2016) used micro- and mini-satellites to examine the diversity within the parasite isolates of the Muguga cocktail vaccine (Hemmink 2014, 2016). It was found that the component stocks contain limited genotypic diversity and, indeed, that two of the three stocks were highly similar. Norling et al. (2015) sequenced the genome of the component stocks of the Muguga cocktail vaccine, also reporting a high degree of similarity between the Serengeti-transformed stock and the Muguga isolate.

Despite its limited diversity, studies in Tanzania have shown that the Muguga cocktail vaccine can achieve good levels of protection (Uilenberg et al. 1977) but despite successful use in some areas, it has been observed that it often does not provide protection against buffalo-derived *T. parva* parasites (Radley et al. 1979). Bishop et al. (2015) and Sitt et al. (2015) recently showed that vaccinated cattle introduced into buffalo-only grazed sites had very little or no protection, with most cattle succumbing to *T. parva* infection. There is the concern that the vaccine can release novel *T. parva* genotypes to unvaccinated cattle (De Deken et al. 2007, Oura et al. 2007) and through recombination more virulent genotypes could be generated (Morzaria and Williamson 1999). Genetic diversity in field populations has vitally important implications for the epidemiology and control of ECF (Muwanika et al. 2016).

The overall aim of this chapter is, therefore, to assess genotypic diversity of *T. parva* populations in cattle and buffalo from the study area where these two species can interact. Antigenic diversity will be examined by using *T. parva*

positive samples from the SENAPA study area described in Chapter 2, as well as several other sample sets from the SENAPA, and applying to them the genotyping pipeline established in Chapter 3.

4.2 Materials and Methods

4.2.1 Study area and samples

As introduced in Chapter 2, the study area is the northern border of the Serengeti National Park (SENAPA), an unfenced boundary where livestock can interact with wildlife.

Cattle samples were from several studies within this study area (Figure 4.1):

1. Cross-sectional cattle survey 2016:
   A cross-sectional survey of the SENAPA study area in 2016, as described in Chapter 2 (n = 770).
2. Cross-sectional cattle survey 2011:
   In 2011 a cross-sectional livestock survey was carried out in a similar study area to the cross-sectional survey of 2016, which sampled cattle from 6 herds to establish prevalence and seroprevalence of FMD in a different project (n = 199).
3. Longitudinal herds:
   As previously described, samples had been collected since 2013, initially in a study investigating Foot and Mouth Disease virus (FMDV) and subsequently in projects analysing animal African trypanosomiasis (AAT); 23 herds were sampled approximately every six to nine months. These samples came from identifiable cattle (individually ear-tagged) so provided a longitudinal sample set of approximately 4,000 samples. Four herds were selected for use from this sample set that covered three time points – 2013, 2015 and 2017 (n = 432).
4. Clinically ill cattle:
   A further subset of the longitudinal samples were from cattle that were described to be clinically ill at the time of sampling (n = 201) – these

were screened for *T. parva* in the expectation that some may have been clinical ECF cases. It was originally hoped to collect samples from confirmed ECF cases in the field but ultimately this was not possible due to logistical reasons.

Buffalo samples (n = 22) were collected from the SENAPA in 2011 (Figure 4.1) (see Chapter 2).

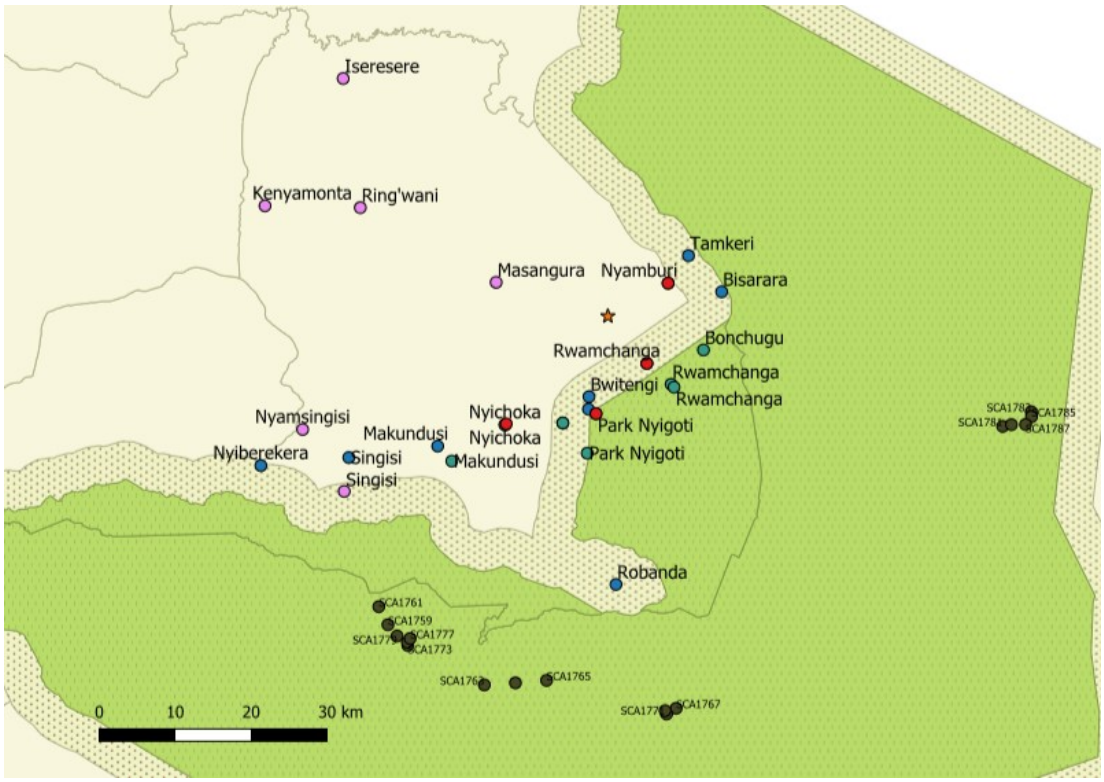

Figure 4.1: Map showing locations of all cattle sample sets as well as buffalo in Serengeti National Park
The eight villages of the 2016 cross-sectional survey are shown in blue. Villages sampled in the 2011 cross-sectional survey are shown in pink. The four selected longitudinal villages are shown in red. Sample locations from animals with clinical signs are shown in green. Buffalo are shown in brown. Boundary shown in dotted beige. Wildlife area shown in green. Field station, Mugumu, is shown as an orange star.

All cattle and buffalo samples were temporarily stored in a cool box during sampling in the field before transferring to storage at -20° at the field station in

Mugumu. Samples were then shipped to The Roslin Institute where they were stored at -80°C or underwent DNA extraction.

4.2.2 DNA extraction

The PAXgene Blood DNA Kit (QIAGEN) was used to isolate genomic DNA from whole blood. Frozen PAXgene blood tubes were thawed at ambient temperature for approximately two hours. Each blood sample (8.5 ml) was mixed by inverting ten times before being transferred to a tube pre-filled with 25 ml of cell lysis buffer BG1. The solution was mixed by inverting five times before centrifuging at 2500 x *g* for five minutes. Supernatant was carefully discarded to leave a pellet before the addition of 5 ml Buffer BG2. The pellet was washed by vortexing for five seconds and then centrifuged at 2500 x *g* for three minutes. Supernatant was removed and 5ml of digestion buffer BG3, including reconstituted PreAnalytiX Protease, was added before vortexing at high speed for 20 seconds to dissolve the pellet. The tube was incubated in a water bath at 65°C for 10 minutes, with the sample colour changing from light red to light green, indicating protein digestion. Samples were vortexed for five seconds before the addition of 5 ml isopropanol (100%) and mixed by inverting 20 times to precipitate the DNA. Tubes were centrifuged at 2500 x *g* for three minutes, the supernatant removed and the tubes inverted on absorbent paper for one minute in order to remove maximal isopropanol. A volume of 5 ml of ethanol (70%) was added to the tube and vortexed for one second before centrifuging at 2500 x *g* for three minutes. Supernatant was removed and the tubes inverted on absorbent paper for five minutes before dabbing the tube to further remove ethanol and leaving inverted for another five minutes in order to dry the pellet. A 1 ml volume of resuspension buffer, BG4, was added and the tube incubated in a water bath at 65°C for one hour, followed by overnight incubation at ambient temperature to dissolve DNA.

4.2.3 Diagnosing *T. parva* in field sample sets

The diagnostic nested p104 PCR assay, described in Chapter 3, was used to screen all the sample sets for prevalence of *T. parva*. A total of 1,602 cattle samples were screened, with 126 being positive for *T. parva* (7.87%). All 22 buffalo samples were positive for *T. parva* (100%). These *T. parva* positive samples (Table 4:1) formed the foundation of the genotypic diversity analysis.

Table 4:1: *T. parva* prevalence in each sample set

Sample set	Samples screened (n)	*T. parva* positive (n)	Prevalence (%)
2016 cross-sectional cattle	770	39	5.06
2011 cross-sectional cattle	199	57	28.64
Longitudinal cattle	432	24	5.56
Clinically ill cattle	201	6	2.99
Buffalo	22	22	100

The species-specific 18S PCR was used as a robust check of a subset of p104 positive samples to assess whether the samples contained only *T.parva* or were co-infected with *T*. sp. (buffalo). Sample size calculation for 'freedom from disease' was calculated expecting 0.00% but allowing for 5.00% error rate, with 95% probability of detecting *T*. sp. (buffalo), and the sample size required to be tested at these thresholds was calculated to be 47. Therefore, 47 p104 positive cattle samples were randomly selected and were screened with the species-specific 18S PCR, resulting in no amplification of *T*. sp. (buffalo). All 22 buffalo samples (100% *T. parva* positive) were screened with the species-specific18S PCR and 18 samples were (weakly) positive for *T.* sp. (buffalo).

4.2.4 Antigen gene selection

The work in Chapter 3 investigated multiple target antigen genes with the aim of amplifying full-length, or near full-length sequences. The preliminary results of the genotyping pipeline demonstrated that the *T. parva* population was similar in cattle and buffalo, albeit in a small number of cattle and buffalo (n =

2 each) in a context where cattle were exposed to high challenge of buffalo-derived parasites, with the majority of sequence clusters shared across the species. Tp1, used as the 'control' antigen gene based on existing prior knowledge (Pelle et al. 2011), did demonstrate greater diversity of the *T. parva* population detected in buffalo compared to that in the cattle samples, and the results for N60 also showed this, to a lesser degree. Antigen gene A14 showed very little diversity and so it was decided not to take this gene forward in this study but it was instead replaced by Tp4 which had also been successfully amplified, as described in Chapter 3, section 3.3.4. Due to frequent cross-reaction with *T*. sp. (buffalo) for the multiple genes investigated, Tp4 was the only other antigen gene for which specific primers were successfully designed and so this gene was selected to take forward in this study. Primer sequences and cycling conditions are as they were described in Chapter 3, with the addition of barcodes to the second round primers (Table 4:2).

Table 4:2: Primers and PCR cycling conditions for candidate antigens

Gene	Primers	Cycling conditions
Tp1		
Round 1	F GCTACGCGGAAATCTAGGCT	98°C for 30 s, 40 cycles (98°C for 10 s, 56°C for 20 s, 72°C for 2.5 min), 72°C for 2 min
	R CATCGTTTGCCAGCACTATGA	
Round 2	F **CACATATCAGAGTGCG**AGGGTCAAAAAAGTTTTATTA	98°C for 30 s, 37 cycles (98°C for 10 s, 54°C for 20 s, 72°C for 1.5 min), 72°C for 2 min
	R **CGCACTCTGATATGTG**TTAATTTTTGAGGTAAATTTTG	
Tp4		
Round 1	F ATACATCCCAAGGCCAAGCT	98°C for 30 s, 40 cycles (98°C for 10 s, 58°C for 20 s, 72°C for 2.5 min), 72°C for 2 min
	R GGAAGGGGTTGGATAGTGCT	
Round 2	F **ACACACAGACTGTGAG**TTACTCATCCTGCCGCTTCT	98°C for 30 s, 30 cycles (98°C for 10 s, 64°C for 20 s, 72°C for 2 min), 72°C for 2 min
	R **CTCACAGTCTGTGTGT**TGACCTCCACCTCTCAACAC	
N60		
Round 1	F TGATCTACAAGCTCGGTGGA	98°C for 30 s, 35 cycles (98°C for 10 s, 68°C for 20 s, 72°C for 1.5 min), 72°C for 2 min
	R GCGGGTATTCTGTGAAGGTC	
Round 2	R **ACACATCTCGTGAGAG**AGACATGGGAAAGGGAAGCT	98°C for 30 s, 32 cycles (98°C for 10 s, 56°C for 20 s, 72°C for 1.5 min), 72°C for 2 min
	F **CTCTCACGAGATGTGT**CCTCCAGTGTCTTTCCGGTA	

Example second round barcode sequences shown in bold

4.2.5 Sample preparation

All p104 positive samples were taken forward for antigen gene amplification (126 cattle, 22 buffalo; Table 4:1). PCR amplicons were generated for antigen genes Tp1, Tp4 and N60. Due to time constraints only data for Tp1 and N60 was analysed fully, and therefore hereafter description and discussion of data will be restricted to these genes. It was not possible to amplify every antigen gene in every DNA sample, most likely due to a combination of low levels of *T. parva* infection in most field samples and varying primer efficiency (Table 4:3).

Table 4:3: Samples resulting in positive amplicon generation

Gene	Cattle*	Buffalo+
Tp1	34	12
N60	79	22

*from a total of 126 p104 nPCR positive cattle
+from a total of 22 p104 nPCR positive buffalo

Nested PCR was used as described previously (Section 3.4.2). For the second round PCR a unique barcode sequence was added to the 5' end of second round primers for each sample to be amplified (Eurofins Scientific); individual amplicons were then combined into a single pool, per antigen gene, for multiplexed sequencing; the addition of barcodes enabled identification of sequences deriving from each amplicon during bioinformatic analysis.

In addition to the field samples, amplicons derived from the reference stock, *T. parva* Muguga (TpM), were also included in each pool as a control. As TpM is a clonal *T. parva* lineage, there should be a single allele with no sequence variation.

In order to be able to assess depth of sequencing and as a technical control, every sample for Tp1 was submitted for sequencing twice, with a further barcode differentiating these batches (referred to as barcode A or barcode B). This step would ensure that if the same collection of variants were present with both A and B barcodes, then the selected depth detected all diversity, but if

sequences were different between A and B, then the full selected depth of sequencing was not sufficient to detect all variants present. Due to time constraints, this step has not yet been analysed and I will not present analysis of these data, but this did impact upon the number of samples that were submitted per SMRT cell.

In this study the amplicons were sequenced on a PacBio Sequel machine, with ~8 fold greater throughput compared to the RS II machine (PacBio 2015-2019). From the trial of the pipeline using RSII, described in Chapter 3, there were approximately 2,000 raw reads generated per amplicon after quality filters were applied. The increased throughput of the Sequel meant that up to 96 samples could be multiplexed and sequenced per SMRT cell, resulting in the expected generation of around 2,000 reads per sample after similar data filtering steps. Factoring in the added step of barcoding each sample twice (A or B) for Tp1, this then allowed up to 48 samples to be sequenced. For N60, for which there was no additional barcode A/B step, up to 96 samples could be sequenced. Amplicons derived from TpM were also included in each antigen pool.

Thus all 46 samples that had generated amplicons for Tp1 (34 cattle, 12 buffalo), along with TpM, were used to generate amplicons for sequencing (n = 94). For N60, 94 samples that had generated amplicons were randomly selected (74 cattle, 21 buffalo), along with TpM (with no duplication) (n = 95) (Table 4:4).

Table 4:4: Number of amplicons from each sample set

	Number of samples/amplicons	
Sample set	**Tp1***	**N60**
2016 cross-sectional cattle	14	25
2011 cross-sectional cattle	12	32
Clinically ill cattle	2	1
2013 longitudinal cattle	2	5
2015 longitudinal cattle	0	5
2017 longitudinal cattle	4	5
Buffalo	12	21
TpM (control)	1	1
Total	**47 x 2 = 94**	**95**

*Tp1 DNA was sampled twice i.e. two amplicons were generated per sample

Qubit fluorimetry was used to measure DNA concentration in order to calculate the volume of each product required in the pooled amplicons to achieve equal representation. Equimolar quantities of PCR products were pooled and Qubit fluorimetry used to verify final DNA concentration of each pool. DNA purity was assessed with NanoDrop spectrophotometry (Thermo Fisher Scientific) and Agilent 2200 TapeStation System (Agilent Technologies) was used to assess DNA integrity and size, using Agilent D5000 Screen Tape System. Ethanol precipitation was carried out to further purify and concentrate the pooled sample, meeting sequencing centre requirements of 3-5 µg at 50 ng/µl per amplicon sample for library preparation. The pooled amplicons (3 µg at 61 ng/µl of Tp1, and 3.3 µg at 65.6 ng/µl of N60) were submitted to Edinburgh Genomics (University of Edinburgh) for sequencing. The two amplicon pools - Tp1 and N60 - were each sequenced on a single SMRT cell on a PacBio Sequel machine (Edinburgh Genomics).

4.3 Bioinformatic methodology

Bioinformatics data handling in sections 4.3.1 and 4.3.2 was carried out by Siddharth Jayaraman (bioinformatician in L. Morrison group); the decision making at each step was done in discussion with myself. As the clustering

analysis carried out in Chapter 3 only resulted in a proxy for diversity, data filtering and analysis in this chapter was changed in order to result in sequence-level data for alleles, and therefore allow assessment of actual gene allele diversity.

Raw data were received as subreads and the following steps were applied. PacBio tools (Bioconda 2019) was used to create circular consensus sequence (CCS) reads (pbccs), to demultiplex (lima) and to create FASTQ format files (bam2fastq). A Matlab script (Jayaraman 2019c) was used for pre-filtering and filtering steps for Tp1 data, and a separate Matlab script (Jayaraman 2019d) was used for N60 data.

4.3.1 Data quality filtering steps

A series of filtering steps were applied to the read data using Matlab scripts for Tp1 (Jayaraman 2019c) and N60 (Jayaraman 2019d).

1. Reads were filtered on the basis of length; reads that fell outside the expected amplicon lengths are likely to be due to fragmented inserts, primer dimers and sequencing artefacts, and therefore a read length filter threshold was set at the mean ± 1SD, based on the observed read length distribution.
2. The 5' and 3' primer sequence information was used to filter out non-full length reads, and all full length reads were converted to be in 5' to 3' orientation.
3. To reduce PacBio sequencing error and retain only high quality reads for diversity estimation, a high 'minimum number of passes' threshold was set based on the observed distribution of full passes for each sample (mean -1SD).

Data that had passed the steps of filtering applied were taken forward.

4.3.2 Analysis of reads post-filtering

The following series of analyses were carried out on the filtered reads:

1. To estimate the allelic diversity in the sequenced reads collected across all samples, first duplicate reads were removed and each read was then pairwise aligned to the respective gene reference sequence using global alignment ('nwalign' function in Matlab).
2. Alternate bases (SNP/INDEL) were identified at each locus along the length of the amplicon, using the alignment information. For each antigen sequenced, a minimum cut-off for SNP/INDEL frequency was set for each dataset to focus on the major variant loci, which were then used as a marker for characterising gene allele distribution. Setting a higher threshold for SNP/INDEL frequency in each case filtered out low level alleles, variations which cannot be reliably differentiated because of background PacBio sequencing errors (Jayaraman et al. 2019) and this also helped in avoiding singleton allele clusters in the next step (alleles supported by only one sequenced read).
3. The marker SNP/INDEL for a given antigen was then used to identify the alleles, by clustering the reads at 100% marker loci identity in order to establish a plausible allele count for each antigen across the sequenced samples.
4. In order to identify particular epitope sequences for each allele, a pseudo-reference was constructed for each allele by substituting the variant alternate base into the reference antigen gene at the variable loci. The amino acid sequence was then translated to identify the epitope sequence.

4.3.3 Gene allele filter

Two levels of filtering were carried out on the allele identified in the dataset in order to establish robust identification of alleles;

1. An allele was required to be present once in at least two individuals to be considered real.
2. An allele was required to be present twice in at least two individuals to be considered real (more stringent filter).

4.3.4 Phylogenetic analysis

Phylogenetic (neighbour-joining) trees were created to examine the allelic diversity and genetic relationships in buffalo- and cattle-derived *T. parva* populations. Gene orthologue sequences from *T. annulata* (TA17450 strain Ankara) were used as an outgroup to root the tree. Bootstrap values of greater than 50% were considered significant (500 replicates). MEGA7 software (Kumar et al. 2016) was used to create the phylogenetic trees.

To further assess evolutionary relationships between alleles, phylogenetic networks were created using SplitsTree4 software (Huson and Bryant 2006).

Heatmaps were generated in R (R Core Team 2018) to illustrate the distribution of alleles in buffalo- and cattle-derived *T. parva* populations.

Nucleotide diversity (pi) and average number of nucleotide differences between alleles (k) were calculated for Tp1 in the buffalo and cattle populations, using PGEToolbox (Cai 2008) in a Matlab script (Jayaraman 2019e).

4.4 Bioinformatics Results

4.4.1 Tp1

A total of 404,758 reads were received; 282,933 for cattle, 113,234 for buffalo and 8,591 for TpM. PacBio Sequel can produce up to 500,000 reads per SMRT cell (PacBio 2015-2019).

Read lengths of pooled inserts showed a peak of expected size, 1618 bp (Figure 4.2).

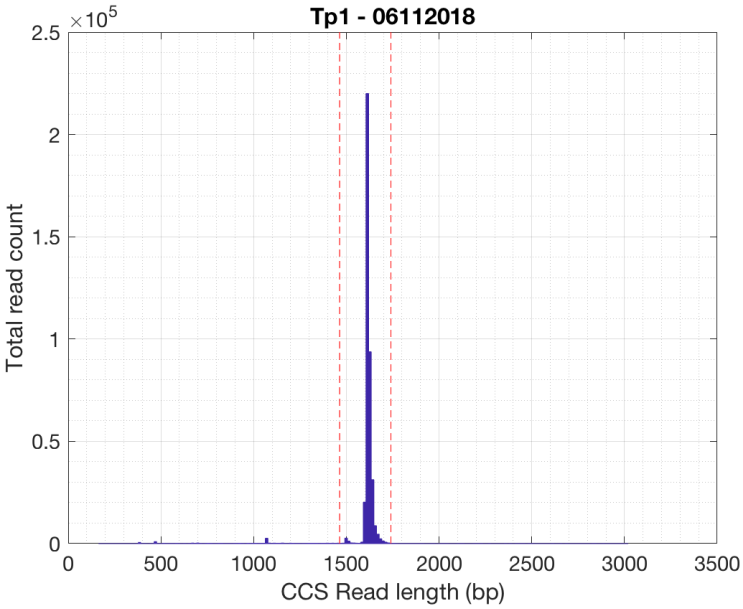

Figure 4.2: Read lengths, including the mean, for Tp1.
Red dashed lines show ± 1SD from the mean.

At the threshold of 15 full passes (mean -1SD), the mean read quality was Phred 90 (equivalent to >99.99% base call accuracy) (Figure 4.3).

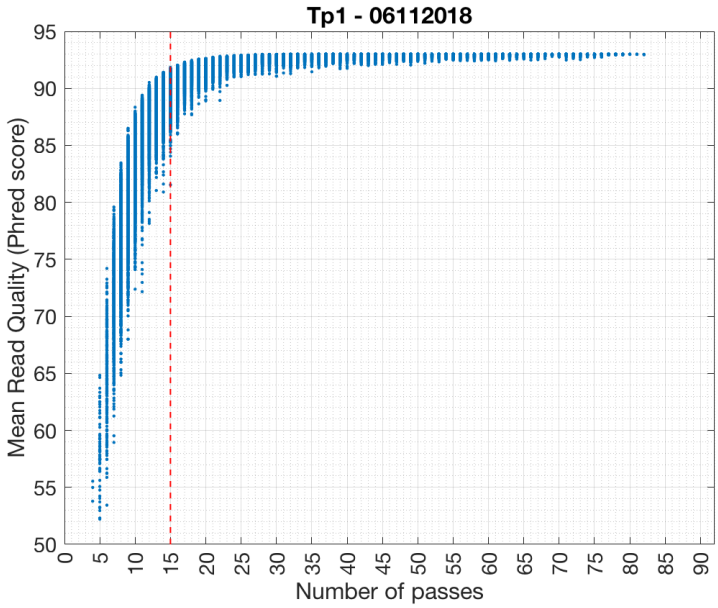

Figure 4.3: Number of passes and associated mean read quality for Tp1
Mean quality score is shown for every read. Red dashed line represents selected threshold level for filtered reads (15 passes)

Post-filtering data were obtained for all samples except one cattle sample (Table 4:5).

Table 4:5: Tp1 read counts for each sample group, before and after quality filtering

Sample Group	Sample Count	CCS Reads	Full length / Insert size / Minimum number of passes filter
TpM	1	8,591	3,400
Buffalo	12	113,234	19,168
Cross-sectional 2016 cattle	14	116, 347	44,564
Cross-sectional 2011 cattle	12	90,029	34,310
Longitudinal cattle	6	55,382	21,514
Clinically ill cattle	2	21,175	8,573
Total	**47**	**404,758**	**131,529**

Read counts, pre- and post-filtering for all individual samples are shown in Appendix E.1

After all filtering steps, the total reads were 131,529 (from a starting number of 404,758). Of these, 108,961 reads (from a pre-filtering total of 282,933) derived from the 34 cattle samples (average of 3,205 reads per cattle sample). 19,168 reads (from a pre-filtering total of 113,234) derived from the 12 buffalo samples (average of 1,597 reads per buffalo sample). 3,400 TpM reads were present after filtering (from 8,591). Therefore, on average twice as many reads were generated in cattle samples compared to buffalo samples. This is not solely due to samples being stored for a longer time before being processed (2011 cattle samples averaged 2,859 reads per sample). This may reflect a lower parasite quantity in the buffalo, although this is difficult to fully explore with only DNA available. In summary, after filtering steps, an overall average of 2,798 reads per sample resulted, which form the basis for analysing the allelic diversity in the *T. parva* population.

4.4.2 Tp1 alleles

To quantify the major alleles observed across all 47 samples from buffalo and cattle, alternate base variants (SNP/INDEL) were identified at each locus over the length of the Tp1 gene after removing the duplicate sequences from the quality filtered PacBio reads. At this stage of the analysis a threshold of >0.1 SNP/INDEL frequency (Figure 4.4) was chosen to focus only on the major alleles, which resulted in mismatches and deletions identified at 93 loci, and insertions were found at 4 loci, i.e. a total of 97 major variable loci in Tp1 (6% of reference length).

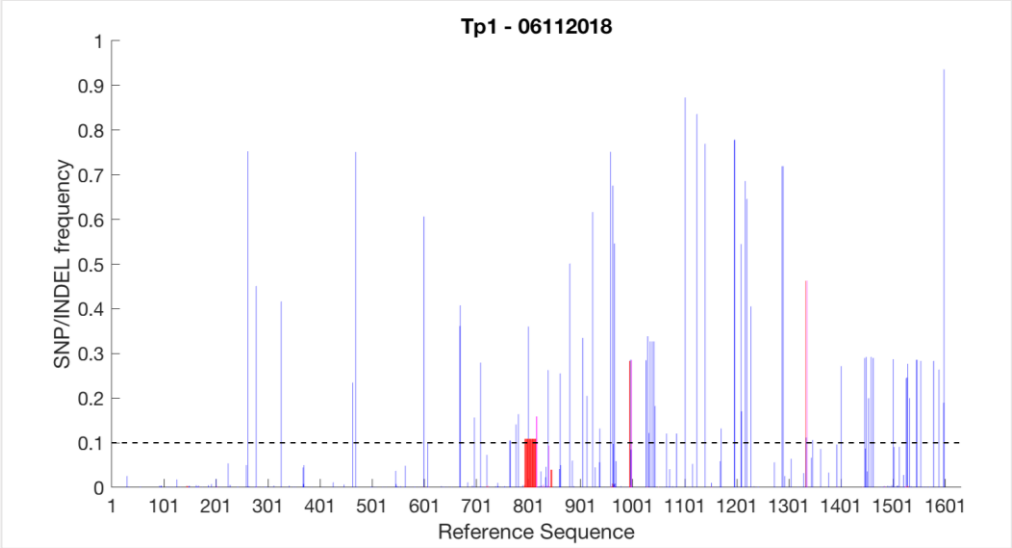

Figure 4.4: SNP/INDEL frequency across the length of the Tp1 Amplicon
The mean calculated SNP/INDEL frequency is shown at each base location for the Tp1 gene. Blue bars represent SNPs and red bars represents INDELs. The horizontal dotted line represents the threshold of 0.1, used to identify major variant loci.

For Tp1 a total of 427 gene alleles were reported but with filtering applied to avoid the presence of alleles represented by single reads, this allele count was reduced, indicating many initially identified alleles were represented by a single read;

1. Level 1 filter required an allele to be present once in at least two individuals; this resulted in 208 alleles, including the reference TpM allele. Read count after this filter was 123,045.
2. Level 2 filter required an allele to be present twice in at least two individuals; this resulted in 86 alleles, including TpM. Read count after this filter was 120,701.

The second, more stringent filter was applied, given that the difference in read number between the two filters was only 2,344 reads, and therefore resolution was not compromised in terms of loss of reads.

Before gene filtering, two alleles were identified as deriving from the TpM control – one allele consisted of just three reads - two reads for TpM and a single cattle read. The other allele consisted of 14,912 reads – 3,396 reads for TpM, 262 buffalo reads and 11,254 cattle reads. The most parsimonious explanation for the former 'allele' is that it derives from PacBio sequencing error. However, this does provide an indication of the level of PacBio introduced error that may be present in the overall data. Ultimately this allele was filtered out by the requirement for at least two reads present in at least two individuals.

Filtered alleles were named in numerical order in relation to their genetic pairwise distance to TpM, with TpM having a distance of 0 (Table 4:6). All alleles were present in at least two individuals, given the filter requirement described.

Table 4:6: Read count distribution per Tp1 allele

Allele ID	Pairwise distance*	Cattle reads	Buffalo reads	TpM reads	Total reads
TpM	0	11,254	262	3396	14,912
Tp1_01	0.001226281	6,670	0	0	6,670
Tp1_02	0.006147297	6,932	0	0	6,932
Tp1_03	0.007381545	73	0	0	73
Tp1_04	0.007381545	19,338	22	0	19,360
Tp1_05	0.008617397	9	0	0	9
Tp1_06	0.009235927	28	0	0	28
Tp1_07	0.010474197	18	0	0	18

Allele ID	Pairwise distance*	Cattle reads	Buffalo reads	TpM reads	Total reads
Tp1_08	0.011714082	11	0	0	11
Tp1_09	0.011714082	9	0	0	9
Tp1_10	0.011714082	6	0	0	6
Tp1_11	0.012334632	23	0	0	23
Tp1_12	0.012334632	16	0	0	16
Tp1_13	0.013576949	9	0	0	9
Tp1_14	0.014198718	108	0	0	108
Tp1_15	0.014820893	19	0	0	19
Tp1_16	0.014820893	18	0	0	18
Tp1_17	0.014820893	122	0	0	122
Tp1_18	0.015443477	0	96	0	96
Tp1_19	0.015443477	31,117	0	0	31,117
Tp1_20	0.016066469	6	0	0	6
Tp1_21	0.016066469	0	93	0	93
Tp1_22	0.016066469	228	0	0	228
Tp1_23	0.016689869	24,375	2	0	24,377
Tp1_24	0.017313679	0	530	0	530
Tp1_25	0.018562529	0	317	0	317
Tp1_26	0.019034406	0	21	0	21
Tp1_27	0.01918757	0	102	0	102
Tp1_28	0.01918757	0	32	0	32
Tp1_29	0.019813023	2	466	0	468
Tp1_30	0.019813023	0	14	0	14
Tp1_31	0.020438887	169	648	0	817
Tp1_32	0.021065165	0	4	0	4
Tp1_33	0.021691855	0	247	0	247
Tp1_34	0.022154133	0	29	0	29
Tp1_35	0.022318959	0	4	0	4
Tp1_36	0.022776961	0	6	0	6
Tp1_37	0.022946478	592	29	0	621
Tp1_38	0.022946478	0	67	0	67
Tp1_39	0.023400198	0	130	0	130
Tp1_40	0.023400198	0	409	0	409
Tp1_41	0.024023844	0	101	0	101
Tp1_42	0.024023844	0	5	0	5
Tp1_43	0.024023844	0	91	0	91
Tp1_44	0.024023844	0	87	0	87
Tp1_45	0.024023844	0	263	0	263
Tp1_46	0.024023844	0	14	0	14
Tp1_47	0.024202759	0	117	0	117
Tp1_48	0.0246479	0	203	0	203

Allele ID	Pairwise distance*	Cattle reads	Buffalo reads	TpM reads	Total reads
Tp1_49	0.0246479	697	493	0	1,190
Tp1_50	0.0246479	0	47	0	47
Tp1_51	0.024831523	474	32	0	506
Tp1_52	0.025272366	304	607	0	911
Tp1_53	0.025272366	0	37	0	37
Tp1_54	0.025897242	0	90	0	90
Tp1_55	0.026090302	1	9	0	10
Tp1_56	0.026090302	0	426	0	426
Tp1_57	0.026090302	0	7	0	7
Tp1_58	0.026720317	0	13	0	13
Tp1_59	0.026720317	0	36	0	36
Tp1_60	0.026720317	459	687	0	1146
Tp1_61	0.026720317	0	35	0	35
Tp1_62	0.026720317	0	185	0	185
Tp1_63	0.026720317	0	35	0	35
Tp1_64	0.02714823	116	250	0	366
Tp1_65	0.027350751	0	14	0	14
Tp1_66	0.027350751	0	24	0	24
Tp1_67	0.027350751	0	11	0	11
Tp1_68	0.027981603	0	993	0	993
Tp1_69	0.027981603	0	297	0	297
Tp1_70	0.028612874	0	6	0	6
Tp1_71	0.032798151	0	168	0	168
Tp1_72	0.033043557	0	160	0	160
Tp1_73	0.033427999	0	89	0	89
Tp1_74	0.033678202	0	11	0	11
Tp1_75	0.034058265	0	9	0	9
Tp1_76	0.034948765	0	2,813	0	2,813
Tp1_77	0.042608038	0	563	0	563
Tp1_78	0.042608038	0	8	0	8
Tp1_79	0.043249107	0	10	0	10
Tp1_80	0.043249107	0	4	0	4
Tp1_81	0.043249107	0	27	0	27
Tp1_82	0.043890609	0	9	0	9
Tp1_83	0.043890609	0	1,423	0	1,423
Tp1_84	0.044532544	0	43	0	43
Tp1_85	0.044532544	0	20	0	20
Total		**103,203**	**14,102**	**3,396**	**120,701**

*Pairwise distance was calculated using ‘seqpdist’ in Matlab, based on Jukes-Cantor method.

Twelve of the 86 alleles were present in both cattle and buffalo, including the reference TpM allele.

The TpM allele was present in 1 buffalo and 9 cattle (21.3% of all individuals). Allele Tp1_23 was present in highest abundance, observed in 20 individuals (42.5%) - 1 buffalo and 19 cattle. The second most abundant allele, Tp1_19, was observed in 17 cattle (36.2%), followed by Tp1_04 which was observed in 13 individuals - 2 buffalo and 11 cattle (27.7%). In contrast, many alleles were present in only two individuals (the lower limit of the allele filter) and these alleles were predominanly in buffalo only. For example, Tp1_75, Tp1_27 and Tp1_53 were each only present in two buffalo (Figure 4.5). A total of 19 allelels were unique to cattle and 55 alleles were unique to buffalo. There was a marked trend in the alleles with the greatest pairwise distance from the TpM sequence tended to be alleles found in buffalo only. In contrast the alleles with the smallest pairwise distance to TpM were cattle alleles, with the exception of a few alleles also seen in buffalo; for example Tp1_04 was present in 11 cattle and 2 buffalo .

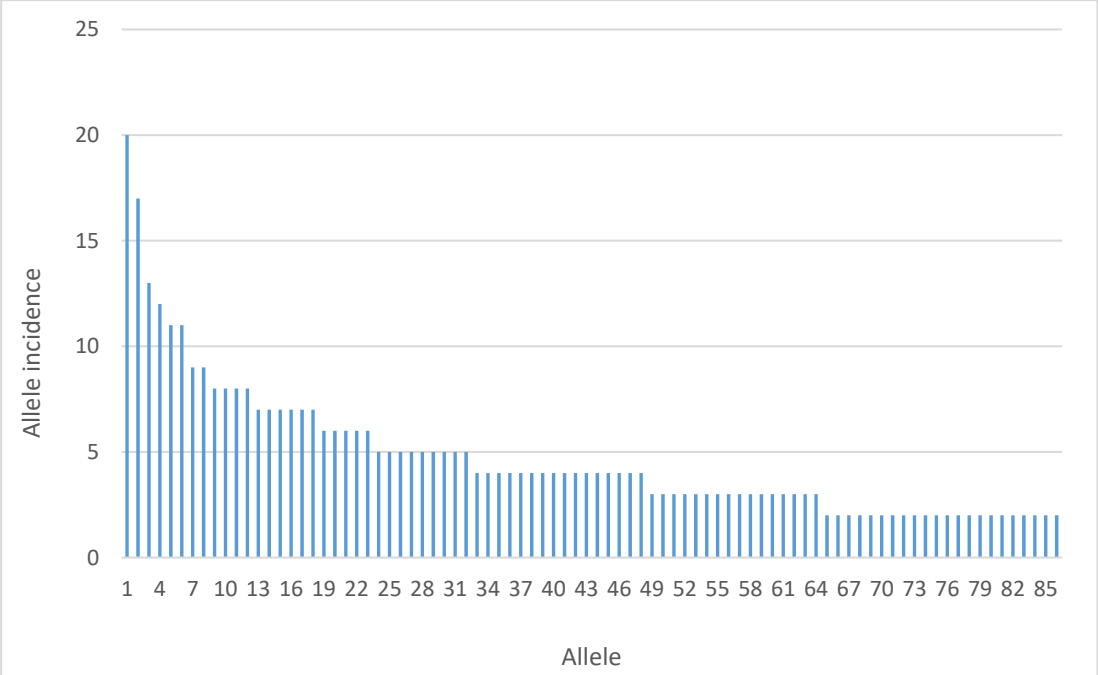

Figure 4.5: Tp1 allele incidence across samples

1=Tp1_23	14=Tp1_12	27=Tp1_17	40=Tp1_45	53=Tp1_59	66=Tp1_77	79=Tp1_82
2=Tp1_19	15=Tp1_40	28=Tp1_74	41=Tp1_37	54=Tp1_65	67=Tp1_15	80=Tp1_57
3=Tp1_04	16=Tp1_64	29=Tp1_13	42=Tp1_06	55=Tp1_05	68=Tp1_36	81=Tp1_63
4=Tp1_49	17=Tp1_01	30=Tp1_22	43=Tp1_79	56=Tp1_67	69=Tp1_42	82=Tp1_58
5=Tp1_52	18=Tp1_44	31=Tp1_62	44=Tp1_51	57=Tp1_26	70=Tp1_21	83=Tp1_03
6=TpM	19=Tp1_72	32=Tp1_11	45=Tp1_38	58=Tp1_28	71=Tp1_16	84=Tp1_75
7=Tp1_56	20=Tp1_47	33=Tp1_39	46=Tp1_54	59=Tp1_30	72=Tp1_69	85=Tp1_27
8=Tp1_50	21=Tp1_84	34=Tp1_78	47=Tp1_08	60=Tp1_71	73=Tp1_66	86=Tp1_53
9=Tp1_83	22=Tp1_25	35=Tp1_33	48=Tp1_14	61=Tp1_10	74=Tp1_70	
10=Tp1_60	23=Tp1_48	36=Tp1_41	49=Tp1_18	62=Tp1_34	75=Tp1_35	
11=Tp1_02	24=Tp1_46	37=Tp1_81	50=Tp1_55	63=Tp1_73	76=Tp1_32	
12=Tp1_31	25=Tp1_29	38=Tp1_09	51=Tp1_76	64=Tp1_20	77=Tp1_80	
13=Tp1_85	26=Tp1_68	39=Tp1_61	52=Tp1_07	65=Tp1_43	78=Tp1_24	

4.4.3 Tp1 CTL epitopes

Using the stringent filtered gene allele dataset, four variants were identified in the previously described sequence that demarcates an epitope recognised by protective cytotoxic CD8 T cells (Table 4:7). The predominant variant circulating in this population was VGYPKVEEII, with 1,383 reads in buffalo and 63,044 reads in cattle, present in 40 (85.1%) individuals. The epitope variant in TpM was VGYPKVKEEML, which was present in 36 (76.6%) individuals. Variant epitopes -MI and -MV were also identified at low levels (10 and 4 individuals, respectively). Variants –ML, -MI, and –II had previously been

reported by Pelle et al. in Kenya (2011). Variant –MV was only reported for the first time in May 2019, and was described from a single cattle isolate from Mara region, Tanzania, with no history of grazing with buffalo (Kerario et al. 2019).

Three of the four epitopes were present in cattle and buffalo (-II, -ML, and -MI). Variant -MV was not present in cattle samples and only found in four buffalo. –II was present in 31 cattle and nine buffalo, -ML was present in 24 cattle and 12 buffalo, and –MI was not found in buffalo but was present in 10 cattle.

Table 4:7: Tp1 epitope variants found in cattle and buffalo samples

Epitope sequence	Number cattle epitope present in	Cattle reads	Number of buffalo epitope present in	Buffalo reads
VGYPKVKEEII	31	63,044	9	1,383
VGYPKVKEEML	24	40,099	12	12,462
VGYPKVKEEMI	10	60	0	0
VGYPKVKEEMV	0	0	4	257

4.4.4 Tp1 allele phylogeny

In order to investigate the relationship between alleles in cattle and buffalo, a phylogenetic tree was constructed (Figure 4.6). The data indicate three broad groups, with a main cattle-derived group (including TpM), a closely related buffalo-derived group with several alleles in this group also being found in cattle, and a third slightly more distinct buffalo-derived group with fewer alleles being present in cattle. However, there was not strong support for substructuring, with all main node bootstrap values less than 50%.

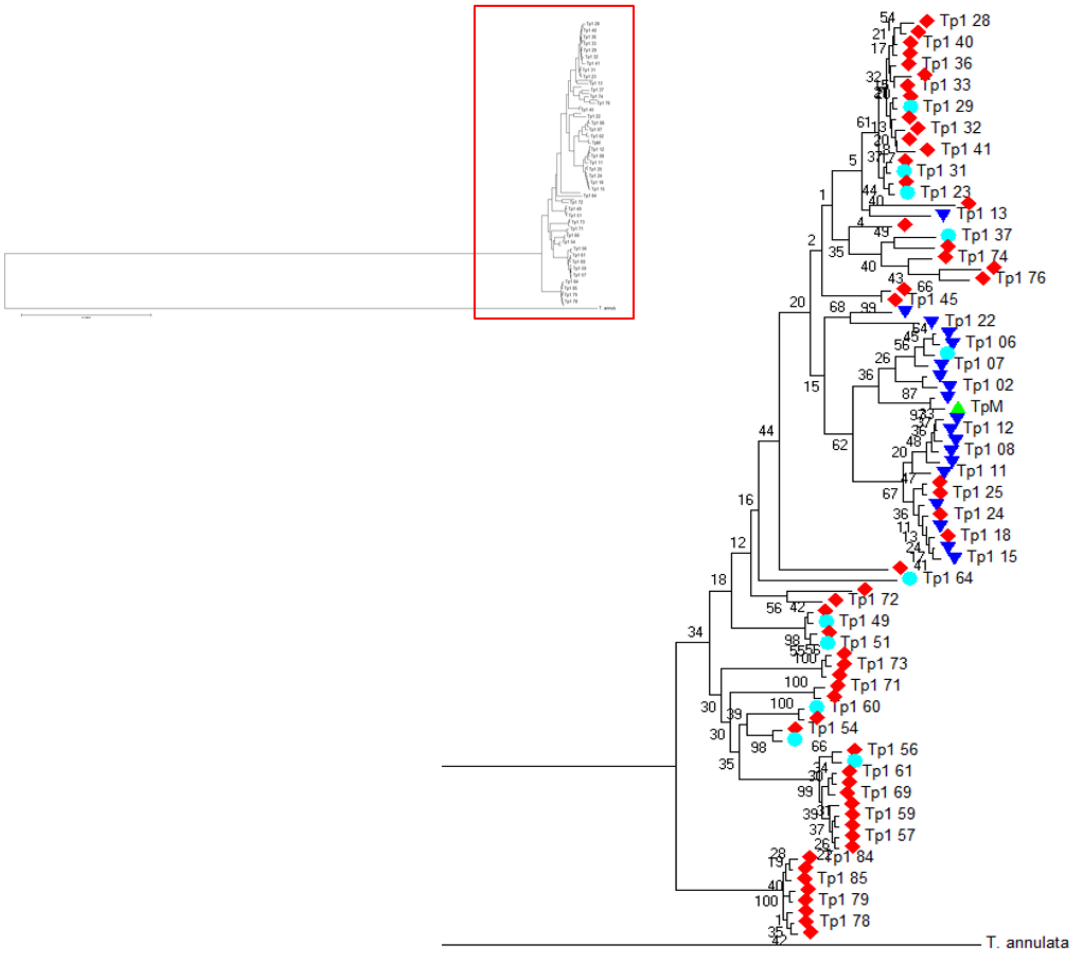

Figure 4.6: Neighbour-joining tree of all 86 Tp1 alleles among cattle and buffalo. *T. annulata* (TA17450 strain Ankara) was used as an outgroup. Alleles unique to cattle-derived *T. parva* are shown in blue, alleles unique to buffalo-derived *T. parva* are shown in red and alleles present in both cattle and buffalo are shown in turquoise. The TpM reference sequence is highlighted in green. For visualisation purposes, the outgroup root is truncated – the full tree is shown in insert (scale bar is 0.05 substitutions per site) with the section of the main figure indicated in the red box. All bootstrap values are shown.

A phylogenetic network was created to further assess evolutionary relationships between alleles (Figure 4.7). The network showed all but one of the cattle-derived alleles clustering in one part, indicating more restricted genetic variability and suggesting the evolutionary introduction of cattle parasite lineages to the population. This cluster also contained the TpM reference sequence. The rest of the network showed the more genetically

diverse buffalo-derived alleles, with ten of the shared alleles distributed amongst them and a single cattle-derived allele (Tp1_13).

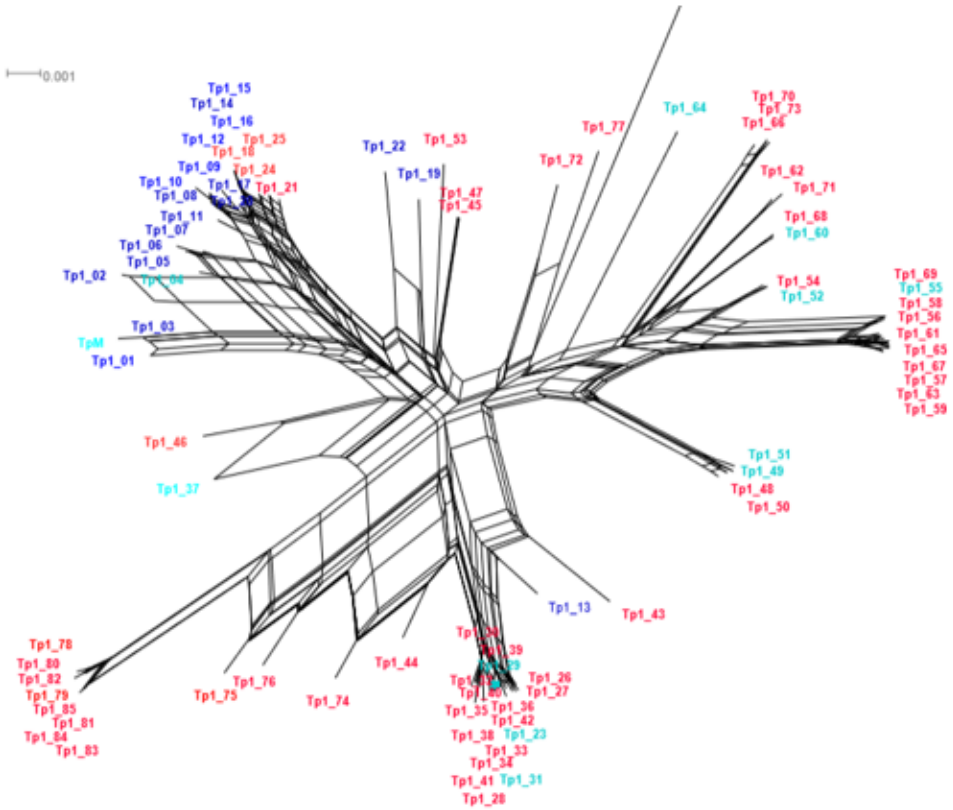

Figure 4.7: Phylogenetic network of all 86 Tp1 alleles among cattle and buffalo. *T. annulata* (TA17450 strain Ankara) is included but for visualisation purposes is truncated and not annotated. Alleles unique to cattle-derived *T. parva* are shown in blue, alleles unique to buffalo-derived *T. parva* are shown in red and alleles present in both cattle and buffalo are shown in turquoise. The scale bar is 0.001 substitutions per site.

A heatmap was generated to assess the relationship of allele distribution in cattle and buffalo (Figure 4.8). As the phylogenetic tree indicated, the data form three clusters; Cluster 1 contains buffalo-derived alleles only, Cluster 2 contains cattle-derived alleles only and Cluster 3 contains predominantly cattle-derived alleles, TpM and four buffalo-derived alleles.

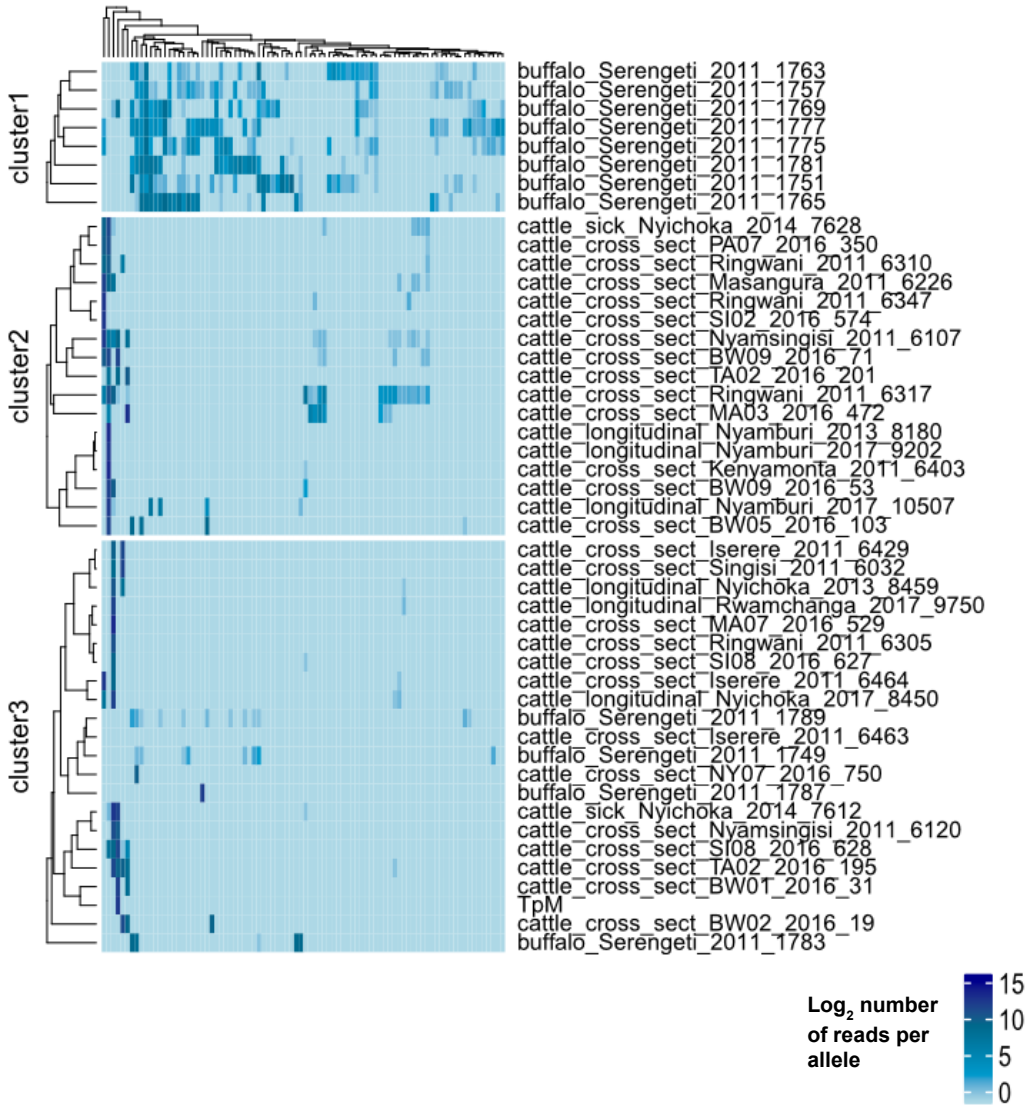

Figure 4.8: Heatmap showing Tp1 allele distribution across samples.
Intensity of colour represents abundance of allele present in individuals. Rows correspond to individual animals and columns correspond to alleles.

The average number of nucleotide differences (k) in the cattle population was 8.900585 and the nucleotide diversity (per site) (pi) was 0.005454. In the buffalo population, the average number of nucleotide differences was 27.096667 and the nucleotide diversity (per site) was 0.016603.

4.4.5 N60

A total of 429,296 reads were received for N60; 246,797 deriving from cattle, 176,195 from buffalo samples and 6304 from TpM.

Read lengths of pooled inserts showed a peak of expected size, 983 bp (Figure 4.9).

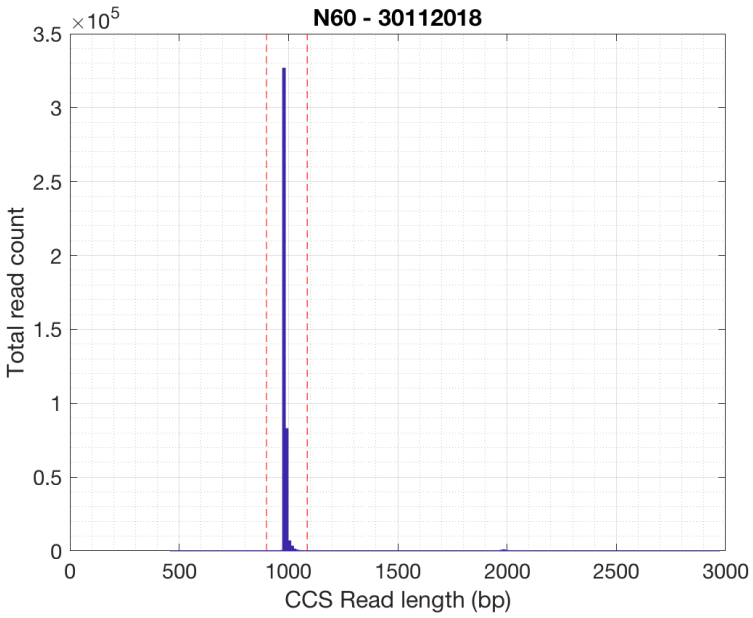

Figure 4.9: Read lengths, including the mean, for N60.
Red dashed lines show ± 1 SD from the mean.

At the threshold of 21 full passes (mean -1SD), the mean read quality was Phred 90 (equivalent to >99.99% base call accuracy) (Figure 4.10).

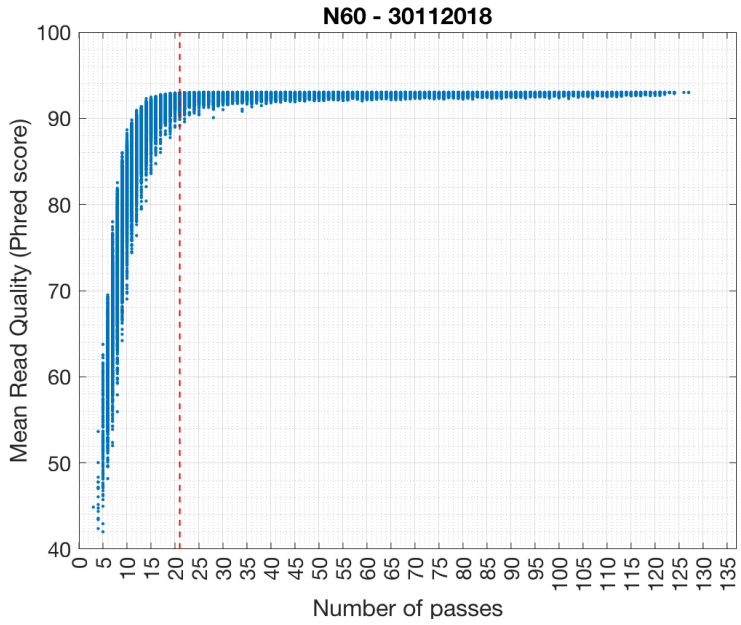

Figure 4.10: Number of passes and associated mean read quality for N60
Mean quality score is shown for every read. Red dashed line represents selected threshold level for filtered reads (21 passes).

Post-filtering data were obtained for 74 individuals (Table 4:8); unfortunately sequence data were not generated for 19 cattle samples and after filtering a further 2 cattle samples were lost.

Table 4:8: N60 read counts for each sample group, before and after read quality filtering

Sample Group	Sample Count	CCS Reads	Full length / Insert Size / Minimum number of passes filtered reads
TpM	1	6,304	3,168
Buffalo	21	176,195	20,532
Cross-sectional 2016 cattle	21	92,964	45,828
Cross-sectional 2011 cattle	22	94,836	46,662
Longitudinal cattle	8	51,104	23,996
Clinically ill cattle	1	7,893	3,604
Total	**74**	**429,296**	**143,790**

Read counts pre- and post-filtering for all individual samples are shown in Appendix E.2

After all filtering steps, the total reads present was 143,790 (from a starting number of 429,296). 120,090 reads (from a pre-filtering total of 246,797) derived from 52 cattle samples (an average of 2,309 reads from the samples for which sequence data was obtained). 20,532 reads (from a pre-filtering total of 176,195) derived from 21 buffalo samples (average read number of 978 reads per buffalo sample), and 3,168 (from 6,303) from TpM. Similar to Tp1, on average twice as many reads were obtained from cattle-derived samples as from buffalo-derived samples. In summary, an average of 1,943 reads per sample was generated – this is lower than Tp1 because for N60, 74 samples were multiplexed into a single flow cell compared to only 47 samples in Tp1. If the increased yield for the Tp1 SMRT cell is taken into account, then the number of reads per sample obtained between runs is broadly similar.

4.4.6 N60 alleles

Similar to Tp1, to quantify the major N60 alleles observed across all 74 isolates from buffalo and cattle, alternate base variants (SNP/INDEL) were identified at each locus over the length of the N60 gene after removing the duplicate sequences from the quality filtered PacBio reads. At this stage of the analysis a threshold of >0.1 SNP/INDEL frequency (Figure 4.11) was chosen to focus only on the major alleles, which resulted in mismatches and deletions identified at 14 loci in the N60 gene (1.4%). No insertions were identified with respect to the reference genome sequence.

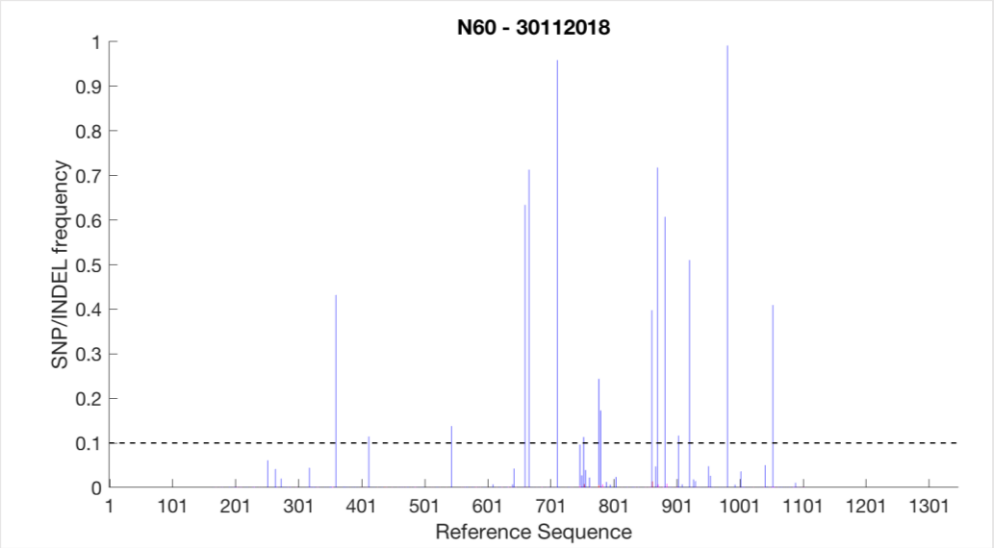

Figure 4.11: SNP/INDEL frequency across the length of the N60 Amplicon.
The mean calculated SNP/INDEL frequency is shown at each base location for the N60 gene. Blue bars represent SNPs and red bars represents INDELs. The horizontal dotted line represents the threshold of 0.1, used to identify major variant loci.

Without taking into account alleles only present once, a total of 578 alleles were identified. However, this allele count was reduced once the following filtering steps were applies, indicating many initially identified alleles were represented by a single read;

1. Level 1 filter required an allele to be present once in at least two individuals; this resulted in 545 alleles, including TpM. Read count after this filter was 143,462.
2. Level 2 filter required an allele to be present twice in at least two individuals; this resulted in 216 alleles, including TpM. Read count after this filter was 142,275.

The second, more stringent filter was applied, given that the difference in read number between the two filters was only 1,187 reads.

In the case of N60, the TpM control was only ever present as a single sequence that matched the TpM reference sequence. Filtered alleles were named in numerical order in relation to their pairwise distance to TpM, with TpM having a distance of 0. All gene alleles were present in at least two

individuals, given the filter requirement described. Of the 216 alleles, 11, including TpM, were present in both cattle and buffalo (Table 4:9).

Table 4:9: Read count distribution per N60 allele

Allele ID	Pairwise distance*	Cattle reads	Buffalo reads	TpM reads	Total reads
TpM	0	3,129	1	3,165	6,295
N60_001	0.000742681	26	0	0	26
N60_002	0.000742681	3	2	0	5
N60_003	0.000742681	41	0	0	41
N60_004	0.000742681	0	119	0	119
N60_005	0.000742681	3	8	0	11
N60_006	0.000742681	54	1	0	55
N60_007	0.001485943	5	0	0	5
N60_008	0.001485943	0	28	0	28
N60_009	0.001485943	8	0	0	8
N60_010	0.001485943	0	49	0	49
N60_011	0.001485943	0	30	0	30
N60_012	0.001485943	0	16	0	16
N60_013	0.001485943	1	14	0	15
N60_014	0.001485943	7	4	0	11
N60_015	0.001485943	0	160	0	160
N60_016	0.001485943	0	121	0	121
N60_017	0.001485943	0	41	0	41
N60_018	0.001485943	9	12	0	21
N60_019	0.001485943	0	335	0	335
N60_020	0.002229786	7	0	0	7
N60_021	0.002229786	0	4,279	0	4,279
N60_022	0.002229786	0	8	0	8
N60_023	0.002229786	1	2,123	0	2,124
N60_024	0.002229786	0	17	0	17
N60_025	0.002229786	0	928	0	928
N60_026	0.002229786	0	13	0	13
N60_027	0.002229786	0	10	0	10
N60_028	0.002229786	0	167	0	167
N60_029	0.002229786	0	10	0	10
N60_030	0.002229786	0	42	0	42
N60_031	0.002229786	0	133	0	133
N60_032	0.002229786	0	12	0	12
N60_033	0.002229786	0	86	0	86
N60_034	0.002229786	4	53	0	57

Allele ID	Pairwise distance*	Cattle reads	Buffalo reads	TpM reads	Total reads
N60_035	0.002229786	0	10	0	10
N60_036	0.002229786	0	13	0	13
N60_037	0.002229786	0	16	0	16
N60_038	0.002229786	0	9	0	9
N60_039	0.002229786	0	10	0	10
N60_040	0.002229786	0	152	0	152
N60_041	0.002229786	0	10	0	10
N60_042	0.002229786	0	8	0	8
N60_043	0.002229786	0	7	0	7
N60_044	0.002229786	0	21	0	21
N60_045	0.002229786	0	16	0	16
N60_046	0.002229786	0	22	0	22
N60_047	0.002974213	0	6	0	6
N60_048	0.002974213	0	9	0	9
N60_049	0.002974213	0	8	0	8
N60_050	0.002974213	4	0	0	4
N60_051	0.002974213	0	26	0	26
N60_052	0.002974213	7	0	0	7
N60_053	0.002974213	0	22	0	22
N60_054	0.002974213	4	0	0	4
N60_055	0.002974213	0	31	0	31
N60_056	0.002974213	0	60	0	60
N60_057	0.002974213	0	31	0	31
N60_058	0.002974213	0	17	0	17
N60_059	0.002974213	0	47	0	47
N60_060	0.002974213	0	31	0	31
N60_061	0.002974213	0	19	0	19
N60_062	0.002974213	0	82	0	82
N60_063	0.002974213	0	16	0	16
N60_064	0.002974213	0	17	0	17
N60_065	0.002974213	0	41	0	41
N60_066	0.002974213	0	9	0	9
N60_067	0.002974213	0	18	0	18
N60_068	0.002974213	0	49	0	49
N60_069	0.002974213	0	8	0	8
N60_070	0.002974213	0	187	0	187
N60_071	0.002974213	0	7	0	7
N60_072	0.002974213	0	18	0	18
N60_073	0.002974213	0	18	0	18
N60_074	0.002974213	0	40	0	40
N60_075	0.002974213	0	13	0	13

Allele ID	Pairwise distance*	Cattle reads	Buffalo reads	TpM reads	Total reads
N60_076	0.002974213	0	16	0	16
N60_077	0.002974213	0	23	0	23
N60_078	0.002974213	0	10	0	10
N60_079	0.002974213	0	27	0	27
N60_080	0.002974213	0	506	0	506
N60_081	0.002974213	0	1,037	0	1,037
N60_082	0.002974213	0	12	0	12
N60_083	0.002974213	0	20	0	20
N60_084	0.002974213	0	93	0	93
N60_085	0.002974213	0	8	0	8
N60_086	0.002974213	12	0	0	12
N60_087	0.002974213	0	26	0	26
N60_088	0.002974213	0	8	0	8
N60_089	0.002974213	0	19	0	19
N60_090	0.002974213	0	14	0	14
N60_091	0.002974213	0	21	0	21
N60_092	0.002974213	0	12	0	12
N60_093	0.002974213	0	13	0	13
N60_094	0.002974213	0	10	0	10
N60_095	0.003719223	0	192	0	192
N60_096	0.003719223	0	45	0	45
N60_097	0.003719223	0	86	0	86
N60_098	0.003719223	0	25	0	25
N60_099	0.003719223	0	11	0	11
N60_100	0.003719223	0	11	0	11
N60_101	0.003719223	0	40	0	40
N60_102	0.003719223	0	199	0	199
N60_103	0.003719223	0	32	0	32
N60_104	0.003719223	0	31	0	31
N60_105	0.003719223	18	0	0	18
N60_106	0.003719223	0	7	0	7
N60_107	0.003719223	0	11	0	11
N60_108	0.003719223	0	24	0	24
N60_109	0.003719223	0	21	0	21
N60_110	0.003719223	0	33	0	33
N60_111	0.003719223	0	40	0	40
N60_112	0.003719223	0	25	0	25
N60_113	0.003719223	0	307	0	307
N60_114	0.003719223	0	191	0	191
N60_115	0.003719223	0	50	0	50
N60_116	0.003719223	0	9	0	9

Allele ID	Pairwise distance*	Cattle reads	Buffalo reads	TpM reads	Total reads
N60_117	0.003719223	0	24	0	24
N60_118	0.003719223	0	64	0	64
N60_119	0.003719223	0	9	0	9
N60_120	0.003719223	0	16	0	16
N60_121	0.003719223	0	9	0	9
N60_122	0.003719223	0	7	0	7
N60_123	0.003719223	0	24	0	24
N60_124	0.003719223	0	737	0	737
N60_125	0.003719223	0	30	0	30
N60_126	0.003719223	0	162	0	162
N60_127	0.003719223	0	14	0	14
N60_128	0.003719223	0	10	0	10
N60_129	0.003719223	0	16	0	16
N60_130	0.003719223	0	29	0	29
N60_131	0.003719223	0	11	0	11
N60_132	0.003719223	5	0	0	5
N60_133	0.003719223	0	8	0	8
N60_134	0.003719223	0	23	0	23
N60_135	0.003719223	0	15	0	15
N60_136	0.003719223	0	5	0	5
N60_137	0.004464818	19	0	0	19
N60_138	0.004464818	0	11	0	11
N60_139	0.004464818	0	13	0	13
N60_140	0.004464818	70	1	0	71
N60_141	0.004464818	0	43	0	43
N60_142	0.004464818	13	1	0	14
N60_143	0.004464818	0	15	0	15
N60_144	0.004464818	0	11	0	11
N60_145	0.004464818	0	4	0	4
N60_146	0.004464818	0	26	0	26
N60_147	0.004464818	0	9	0	9
N60_148	0.004464818	0	7	0	7
N60_149	0.004464818	0	10	0	10
N60_150	0.004464818	0	9	0	9
N60_151	0.004464818	0	33	0	33
N60_152	0.004464818	0	82	0	82
N60_153	0.004464818	0	36	0	36
N60_154	0.004464818	0	30	0	30
N60_155	0.004464818	0	13	0	13
N60_156	0.004464818	65	0	0	65
N60_157	0.004464818	0	15	0	15

Allele ID	Pairwise distance*	Cattle reads	Buffalo reads	TpM reads	Total reads
N60_158	0.004464818	0	11	0	11
N60_159	0.004464818	0	37	0	37
N60_160	0.004464818	0	21	0	21
N60_161	0.004464818	0	8	0	8
N60_162	0.004464818	0	8	0	8
N60_163	0.004464818	0	115	0	115
N60_164	0.004464818	0	24	0	24
N60_165	0.004464818	0	297	0	297
N60_166	0.004464818	0	1,847	0	1,847
N60_167	0.004464818	0	38	0	38
N60_168	0.004464818	0	27	0	27
N60_169	0.004464818	0	27	0	27
N60_170	0.004464818	0	70	0	70
N60_171	0.004464818	0	107	0	107
N60_172	0.004464818	0	64	0	64
N60_173	0.004464818	0	10	0	10
N60_174	0.004464818	0	7	0	7
N60_175	0.005210999	116,418	0	0	116,418
N60_176	0.005210999	0	26	0	26
N60_177	0.005210999	0	55	0	55
N60_178	0.005210999	0	67	0	67
N60_179	0.005210999	0	14	0	14
N60_180	0.005210999	0	39	0	39
N60_181	0.005210999	0	12	0	12
N60_182	0.005210999	0	36	0	36
N60_183	0.005210999	0	9	0	9
N60_184	0.005210999	0	8	0	8
N60_185	0.005210999	0	8	0	8
N60_186	0.005210999	0	16	0	16
N60_187	0.005210999	0	443	0	443
N60_188	0.005210999	0	44	0	44
N60_189	0.005210999	0	12	0	12
N60_190	0.005210999	0	65	0	65
N60_191	0.005210999	0	14	0	14
N60_192	0.005210999	0	12	0	12
N60_193	0.005210999	0	31	0	31
N60_194	0.005210999	0	9	0	9
N60_195	0.005210999	0	109	0	109
N60_196	0.005210999	0	7	0	7
N60_197	0.005210999	0	9	0	9
N60_198	0.005957767	0	61	0	61

Allele ID	Pairwise distance*	Cattle reads	Buffalo reads	TpM reads	Total reads
N60_199	0.005957767	0	24	0	24
N60_200	0.005957767	0	14	0	14
N60_201	0.005957767	0	12	0	12
N60_202	0.005957767	0	9	0	9
N60_203	0.005957767	0	8	0	8
N60_204	0.005957767	0	15	0	15
N60_205	0.005957767	0	14	0	14
N60_206	0.005957767	0	26	0	26
N60_207	0.005957767	0	50	0	50
N60_208	0.005957767	0	9	0	9
N60_209	0.005957767	0	227	0	227
N60_210	0.005957767	0	11	0	11
N60_211	0.006705121	0	9	0	9
N60_212	0.006705121	0	46	0	46
N60_213	0.006705121	0	5	0	5
N60_214	0.006705121	0	63	0	63
N60_215	0.006705121	0	18	0	18
Total		**119,933**	**19,177**	**3,165**	**142,275**

*Pairwise distance was calculated using 'seqpdist' in Matlab, based on Jukes-Cantor method.

The TpM allele was present in 5 cattle and 1 buffalo (6%). Allele N60_175 was present in greatest abundance, observed in 52 cattle (54% of all individuals). N60_023 was present in 20 buffalo and one cow (22%) and N60_019 was present in 20 buffalo only (21%). In contrast, N60_009 was present in only two cattle, as was N60_007, and N60_145 was present in only two buffalo (2%) (Figure 4.12). A total of 14 alleles were unique to cattle and 191 alleles were unique to buffalo. Eleven alleles, including TpM, were shared across both cattle and buffalo.

Similar to Tp1, the alleles with the greater pairwise distances from TpM were almost all buffalo-derived alleles, with the exception of N60_156 and N60_175 that are cattle-derived alleles. In general, the few alleles that were unique to cattle had small pairwise distances from TpM, such as N60_01, N60_03 and N60_07.

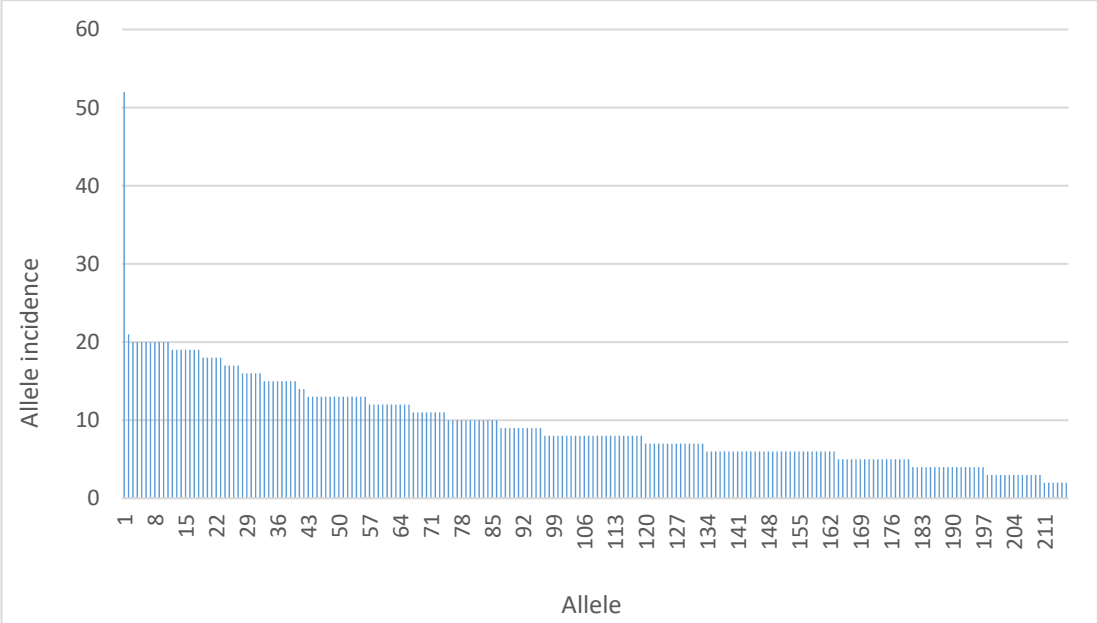

Figure 4.12: N60 allele incidence across samples

1=N60_175	38=N60_046	75=N60_012	112=N60_076	149=N60_097	186=N60_105
2=N60_023	39=N60_057	76=N60_153	113=N60_199	150=N60_048	187=N60_047
3=N60_019	40=N60_028	77=N60_087	114=N60_127	151=N60_042	188=N60_148
4=N60_126	41=N60_068	78=N60_018	115=N60_204	152=N60_192	189=N60_014
5=N60_080	42=N60_110	79=N60_189	116=N60_041	153=N60_072	190=N60_106
6=N60_081	43=N60_178	80=N60_045	117=N60_139	154=N60_194	191=N60_211
7=N60_124	44=N60_182	81=N60_206	118=N60_131	155=N60_069	192=N60_183
8=N60_031	45=N60_065	82=N60_055	119=N60_073	156=N60_202	193=N60_032
9=N60_165	46=N60_187	83=N60_082	120=N60_024	157=N60_112	194=N60_174
10=N60_166	47=N60_172	84=N60_212	121=N60_063	158=N60_061	195=N60_197
11=N60_033	48=N60_016	85=N60_064	122=N60_035	159=N60_119	196=N60_144
12=N60_104	49=N60_209	86=N60_109	123=N60_134	160=N60_029	197=N60_108
13=N60_084	50=N60_160	87=N60_164	124=N60_193	161=N60_006	198=N60_198
14=N60_015	51=N60_074	88=N60_151	125=N60_138	162=N60_078	199=N60_196
15=N60_114	52=N60_207	89=N60_083	126=N60_005	163=N60_173	200=N60_088
16=N60_025	53=N60_103	90=N60_104	127=N60_TpM	164=N60_085	201=N60_020
17=N60_034	54=N60_154	91=N60_053	128=N60_044	165=N60_203	202=N60_143
18=N60_062	55=N60_125	92=N60_058	129=N60_147	166=N60_135	203=N60_094
19=N60_040	56=N60_141	93=N60_170	130=N60_027	167=N60_122	204=N60_132
20= N60_070	57=N60_169	94=N60_117	131=N60_149	168=N60_022	205=N60_052
21= N60_095	58=N60_101	95=N60_140	132=N60_150	169=N60_184	206=N60_213
22=N60_118	59=N60_146	96=N60_157	133=N60_116	170=N60_071	207=N60_003
23=N60_030	60=N60_102	97=N60_089	134=N60_180	171=N60_039	208=N60_136
24=N60_021	61=N60_168	98=N60_137	135=N60_038	172=N60_201	209=N60_205
25=N60_163	62=N60_060	99=N60_200	136=N60_191	173=N60_190	210=N60_086
26=N60_188	63=N60_156	100=N60_100	137=N60_075	174=N60_129	211=N60_002
27=N60_096	64=N60_051	101=N60_179	138=N60_036	175=N60_043	212=N60_050
28=N60_115	65=N60_098	102=N60_037	139=N60_161	176=N60_142	213=N60_054
29=N60_056	66=N60_091	103=N60_092	140=N60_066	177=N60_158	214=N60_009
30=N60_159	67=N60_011	104=N60_155	141=N60_162	178=N60_214	215=N60_007
31=N60_111	68=N60_077	105=N60_093	142=N60_133	179=N60_107	216=N60_145
32=N60_195	69=N60_123	106=N60_067	143=N60_152	180=N60_099	
33=N60_177	70=N60_013	107=N60_176	144=N60_128	181=N60_130	
34=N60_017	71=N60_008	108=N60_215	145=N60_049	182=N60_210	
35=N60_159	72=N60_167	109=N60_181	146=N60_121	183=N60_186	
36=N60_113	73=N60_120	110=N60_026	147=N60_208	184=N60_001	
37=N60_010	74=N60_079	111=N60_090	148=N60_185	185=N60_171	

There were no epitope variants found for N60, as all nucleotide alterations in the gene alleles were synonymous, resulting in no coding changes.

4.4.7 N60 allele phylogeny

A phylogenetic (neighbour-joining) tree was generated to examine the relationship between alleles in cattle- and buffalo-derived *T. parva* populations (Figure 4.13). The tree illustrates cattle alleles appearing more closely related to TpM and grouping together, but this is less clear than in Tp1, and there was not strong support for substructuring with all main bootstrap values less than 50%.

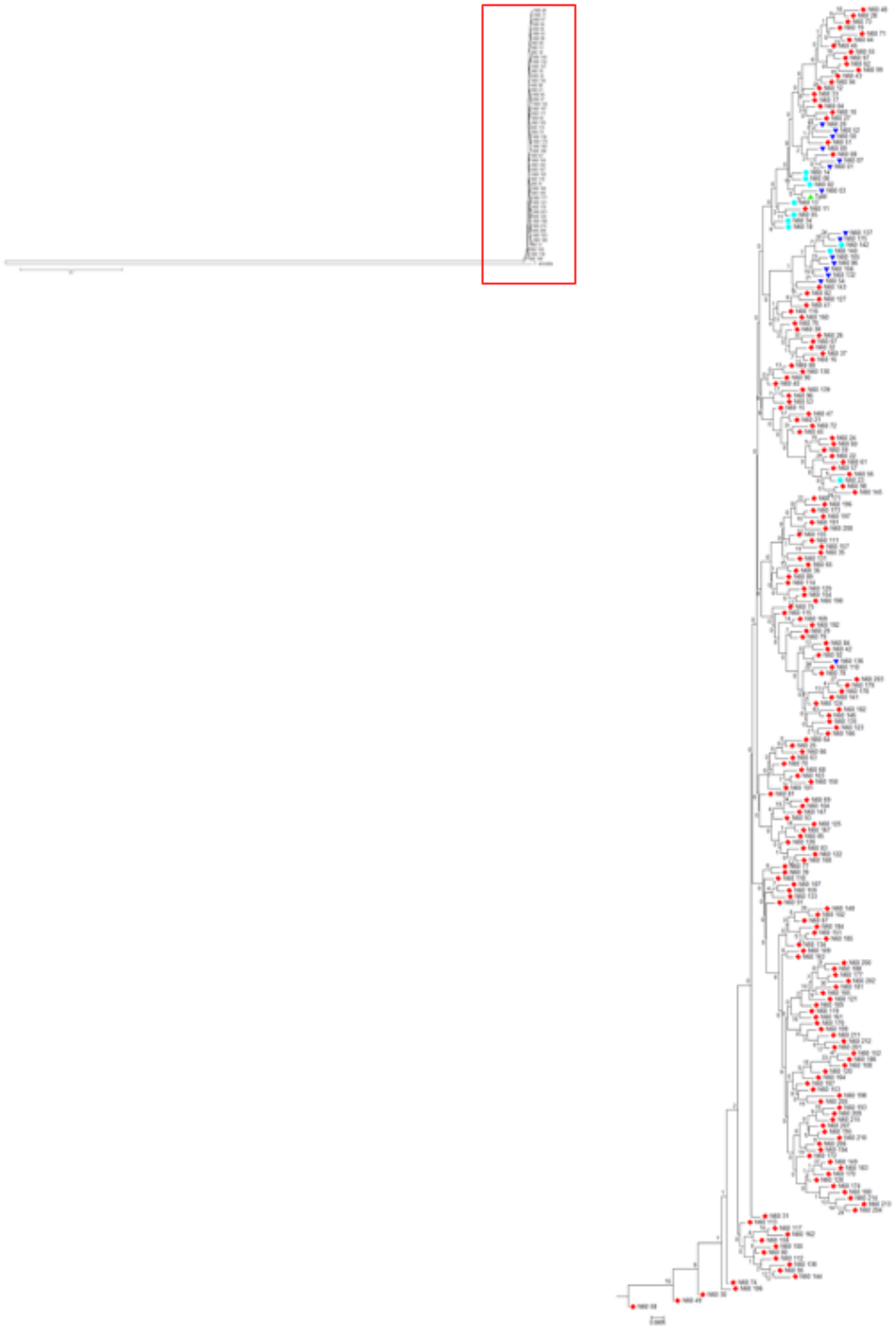

Figure 4.13: Neighbour-joining tree of 216 N60 alleles among cattle and buffalo. *T. annulata* (TA17450 strain Ankara) was used as an outgroup. Alleles unique to cattle-derived *T. parva* are shown in blue, alleles unique to buffalo-derived *T. parva* are shown in red and alleles present in both cattle and buffalo are shown in turquoise. The TpM reference sequence is highlighted in green. For visualisation purposes, the outgroup root is truncated – the full tree is shown in insert (scale bar is 0.1 substitutions per site) with the section of the main figure indicated in the red box (scale bar is 0.0005). All bootstrap values are <50%.

A phylogenetic network was created to further assess evolutionary relationships between N60 alleles (Figure 4.14). The cattle-derived alleles are shown in two clusters relatively close to each other in the network, suggesting the introduction of cattle-derived parasite genetic lineages with some spread within the cattle population. The shared alleles are distributed amongst and between the cattle-derived allele clusters. Buffalo-derived alleles are distributed between the cattle-derived clusters, as well as forming four other clusters and making up the bulk of the network.

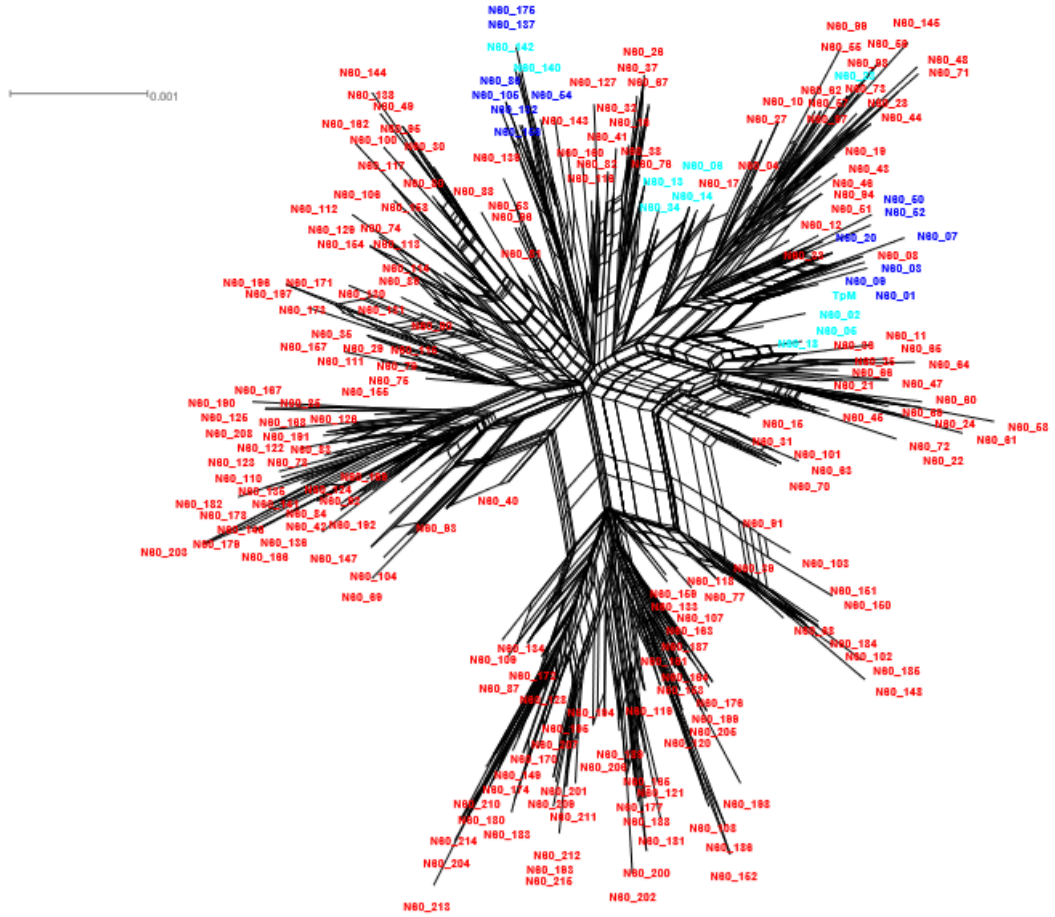

Figure 4.14: Phylogenetic network of all 216 N60 alleles among cattle and buffalo.
Alleles unique to cattle-derived *T. parva* are shown in blue, alleles unique to buffalo-derived *T. parva* are shown in red and alleles present in both cattle and buffalo are shown in turquoise. The scale bar is 0.001 substitutions per site.

A heatmap was generated to assess the relationship of allele distribution in cattle and buffalo (Figure 4.15). Although no actual clusters, there is grouping of buffalo-derived alleles, and grouping of cattle-derived alleles with TpM, as well as one buffalo-derived allele that has 4206 reads in buffalo_1787 and is present in much lower abundance (1-10 reads) in 16 other buffalo. The absence of allele N60_175 from the buffalo samples (present in 52 cattle) also separates the relationship.

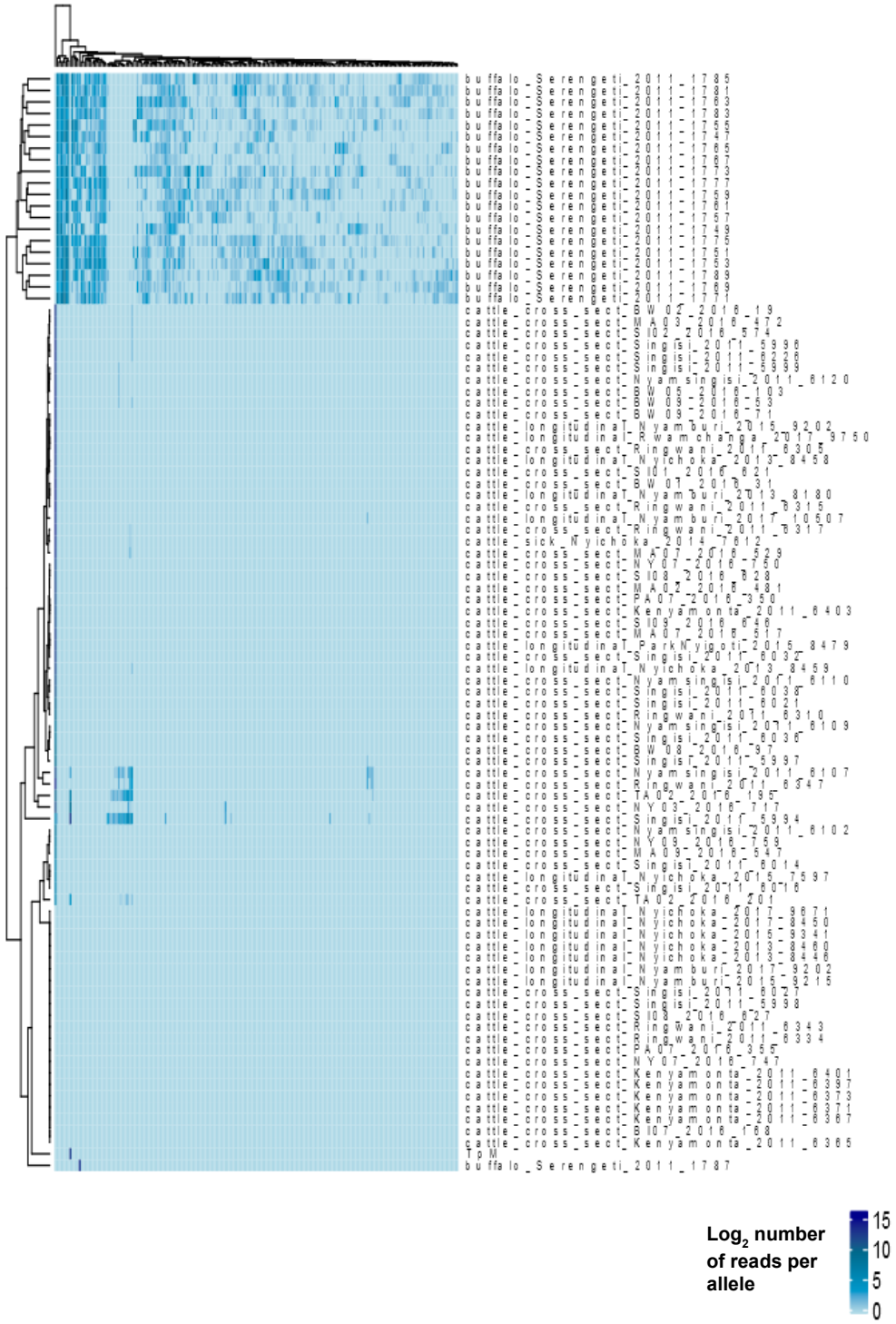

Figure 4.15: Heatmap showing N60 allele distribution across samples. Intensity of colour represents abundance of allele present in individuals. Rows correspond to individual animals and columns correspond to alleles.

To summarise, variable nucleotide loci were identified in both Tp1 (n = 97) and N60 (n = 14); 86 alleles were observed in Tp1, and 216 alleles in N60.

Despite the high number of allelic variants at the nucleotide level, these are based on a much smaller number of variant loci, and the predicted antigenic diversity of N60 was much more limited. All of the nucleotide changes in N60 were synonymous resulting in no coding changes and thus no epitope variants, the identified epitope variant then being identical to TpM. Four epitope variants were observed in Tp1, with two being common (>75% of samples) – the TpM variant being one of these. The predominant variant was VGYPKVEEII. The epitope variant found in the reference stock, TpM, was VGYPKVKEEML.

Most of the alleles in both Tp1 and N60 were unique to buffalo – 55 of 86 in Tp1, and 191 of 216 in N60. A smaller proportion were unique to cattle (19 of 86 in Tp1, 14 of 216 in N60) with the remainder shared in both cattle and buffalo (12 of 86 in Tp1, 11 of 216 in N60).

4.4.8 Discussion

In this study, *T. parva* antigen genes Tp1 and N60 were sequenced and population-level full length or near-full length sequences were analysed in order to assess the genetic and antigenic diversity. A high level of diversity at the nucleotide level was observed in both Tp1 and N60 resulting in 86 allele variants of Tp1 and 216 allele variants of N60. There was greater allelic diversity seen in buffalo for both antigen genes, with a higher number of alleles unique in buffalo than in cattle and only a relatively small proportion unique to cattle or shared in both host species. The results support the findings of previous studies that there is greater *T. parva* parasite heterogeneity in buffalo-derived populations compared to the cattle-derived population (Bishop et al. 1994a, Oura et al. 2011a, Pelle et al. 2011).

The results of Tp1 are in agreement with previous findings of Pelle et al. (2011) where diverse allelic variation was observed in Tp1, with greater diversity in buffalo-derived parasites than cattle-derived, and very recently a study by Kerario et al. (2019) also demonstrated higher levels of Tp1 diversity in buffalo-derived isolates than in cattle-derived. To date, there have been no other studies examining the degree of diversity of antigen N60 and so these findings represent novel data for this antigen.

The alleles in both Tp1 and N60 demonstrated a high degree of relatedness, consistent with the samples all deriving from a largely co-circulating population. There is, however, evidence of substructuring in Tp1, as shown in the heatmap and network analysis, indicating some separation between parasites deriving from cattle and buffalo, with a group of buffalo-derived samples being slightly more genetically distant and being rarely found in cattle. For N60, there is more limited evidence for substructuring in terms of host (although the N60 data had limited resolution given there was only polymorphism at 14 nucleotide positions). As the *T. parva* parasite population is likely to be ancient, with most allelic diversity evolving over evolutionary time, it is unlikely that there has been sufficient passage of time for complete genetic substructuring to be evident based on a one or two loci.

Individual animals often had many alleles present indicating a high level of multiplicity of infection (MOI). Mixed infections are a common occurrence in field samples, often due to the parasite diversity within the tick vector (Elisa et al. 2015), which can be the result of recombination in the tick (Hayashida et al. 2013, Katzer et al. 2011). Hemmink (2014) had previously observed large numbers of alleles in individual buffalos from Kruger, South Africa, and Ol Pejeta, Kenya, using multilocus sequence typing. Studies have shown that ticks usually have only very few infected salivary gland acini, where the sporozoites in each acini have originated from one parasite (Gitau et al. 2000), which would then indicate that extensive allelic diversity within individual animals is the result of multiple tick bites and ensuing infections, and the diversity present in ticks would in turn be defined by whether ticks had previously fed on buffalo or cattle.

The nucleotide polymorphism observed in Tp1 resulted in changes in amino acid residues and 85 amino acid variants were identified. Consequently, four variants were observed in the previously defined epitope recognised by protective CTL responses against Tp1. The predominant variant was VGYPKVEEII. Pelle et al. (2011) and Kerario et al. (2019) found variant –ML in the majority of their isolates, from Kenyan and Tanzanian isolates respectively. The epitope variant in the reference stock, TpM, was VGYPKVKEEML, as observed in other studies (Elisa et al. 2015, Graham et al. 2008, Hemmink et al. 2018, Pelle et al. 2011), providing good validation of the pipeline and extending confidence in the N60 data as well. Antigen variant –MV was only observed in buffalo and has only recently been reported for the first time, but in a single cow from Mara region in northern Tanzania (Kerario et al. 2019). The predominant variant in this study, –II, was the second-most common variant found by Kerario and Pelle, present in cattle and buffalo from diverse regions in Tanzania and from Kenya respectively. Due to the array of geographical origins of the samples and isolates analysed, across all studies, it is difficult to make any inferences about the variation in the Tp1 epitope in regards to location.

Despite the high number of gene alleles observed in N60, there were no non-synonymous variants. No variation at the protein level is an interesting finding and could suggest purifying selection, by the removal of deleterious polymorphisms. Genetic diversity is reduced by purifying selection (Cvijovic et al. 2018) and although the mechanisms are not fully understood, it could be predicted that there is an important role in conserving the amino acid sequence to maintain vital functions for parasite survival. N60 has been identified as being recognised by CD4 T cell responses in immune cattle. As the role of CD4 T cell responses in protection is unclear, it is not known whether such antigens are under immune selection. This study demonstrates that the pipeline created can assess conservation of proteins at the population level, and this could be useful to assess potential vaccine candidates. The use of a conserved gene in an immunisation could lead to broad immunity, protective whether buffalo or cattle-derived.

As previously discussed, buffalo-derived parasites do not differentiate well into the piroplasm state in cattle (Schreuder et al. 1977) and many of the attempts to transmit infection from cattle infected with buffalo-derived parasites have been unsuccessful. The cattle samples were almost all from healthy animals, with the exception of the clinically ill samples, therefore the cattle samples are likely to predominantly represent the carrier state, with circulating piroplasms. The *T. parva* parasites in the cattle samples are, therefore, likely to be representative of the cattle-maintained parasite population. Due to the nature of the blood samples being used, it could be that there are buffalo-derived parasites present in cattle but that they are not being detected in this study, thus creating a founder effect. In order to definitively detect infection with buffalo-derived parasites in cattle, it would be necessary to sample cattle exposed to infected ticks in an area not previously grazed by cattle. A recently published study by Sitt et al. (2019) has reported marked diversity in the sequence of the p67 gene in cattle infected under these conditions, similar to that seen in buffalo. This indicated that the different genotypes found in buffalo are able to establish acute infections in cattle. However, since previous studies had detected only a single allele of p67 in cattle-maintained *T. parva*, these

findings also support the view that these buffalo-derived infections would not be maintained in the cattle population by ticks. The findings of the current study indicate that the parasite populations being detected (presumably at the piroplasm stage for both) are different in cattle and buffalo, with most alleles of each antigen unique to one or other host species and only comparatively few shared between both. However, the presence of alleles shared between cattle and buffalo does indicate that there is mixing of parasite populations. The phylogenetic relationship of both Tp1 and N60 indicate that the parasite population is similar, despite the common separation of alleles by host species. As it is suggested that there is sharing of parasites, there are important implications for use of the ITM vaccine in the area if buffalo isolates differ to those in cattle. The level of interaction between cattle and buffalo, and the ticks that feed on them, in the study area is unknown. However, discussion with farmers in the study area described the grazing of cattle within the protected areas (particularly seasonally when water supplies were limited) so there is clear opportunity for the tick vector to be shared between mammalian hosts, allowing for parasites also to be shared, and the data presented backs this up in that there are not genetically divergent *T. parva* populations in cattle and buffalo.

The TpM reference allele sequence was present in 9 cattle and 1 buffalo for Tp1, and in 5 cattle and 1 buffalo for N60, which is interesting given the significant divergence in time and parasite generations between *T. parva* Muguga being originally isolated in the 1950s and cultured and passaged in the laboratory and the time of sampling for the current samples being analysed (2011-2017). It was assessed to see if there was a pattern within individual samples in which the TpM allele sequence was present, for both antigen genes. The TpM allele was present in a single buffalo for Tp1 and a different single buffalo for N60. However, it was present for both antigen genes in three cattle from two sampling time points: cross-sect_Nyamsingisi_2011_6106, cross-sect_TA02_2016_196 and cross-sect_TA02_2016_201. These data indicate that parasites containing the TpM reference sequence are present in both cattle and buffalo and across the time period of sampling (between 2011

and 2016). Although data in the questionnaire survey in Chapter 2 indicated that none of the farmers in the study area were using the ITM vaccine, it is possible that vaccinated cattle could have been brought into the area and thus introduced the vaccine strain to feeding ticks, which could then in turn infect cattle and buffalo with the vaccine strain. Pelle et al. (2011) also reported the presence of the TpM allele in isolates deriving from Kenya. In the absence of evidence of vaccinated cattle in the area, the widespread presence of these sequences may simply indicate a circulating allele, present at an appreciable level with an as yet unidentified selective pressure that maintains this sequence, and un-related to ITM vaccination. This can only be resolved by a fuller assessment of parasite diversity across the range of *T. parva*.

There were a variety of potential factors that could have influenced the samples sequenced (i.e. host species, time, space, management practice etc.). It would have been interesting to assess how the genotypic composition of *T. parva* parasites changes spatially and temporally. Although samples from different time points and farms contributed to the total samples analysed, the numbers of samples per time or location resulting in successful generation of sequence data were small, and would therefore be under-powered to make inferences about parasite populations at specific areas or times. Although spatiotemporal analysis was not formally carried out, there was no strong trend in allele distribution across space or time evident in the data.

The bioinformatics processing of reads used a stringent filtering approach. There was a large reduction in reads seen from CCS to filtered stage. This loss occurred at the read length filter, suggesting an erroneous event with the primers in the preparation of the samples. There is the potential for considerable sequencing error in the generation of high read numbers, but with the creation of consensus and by using a stringent filtered approach to the sequence data, the rate of error should be considerably reduced. However, there is still the possibility of errors introduced during PCR. Incomplete extension of primers during PCR can result in the generation of artificial chimeras; a partially extended product can anneal to a different template

sequence during subsequent PCR cycles, resulting in hybrid sequences known as artificial chimeras (Meyerhans et al. 1990). Chimera formation during PCR can create false variants (Smyth et al. 2010) which could inflate the allele count. Therefore, an additional step could be to add a chimera filter during the read filtering. It should also be acknowledged that low abundance sequences may not be well detected; although this is not an error, it can create a bias and result in inaccurate reflection of the template DNA composition (Schloss and Westcott 2011). Also we focused only on the major alleles in the population by using a higher threshold for SNP/INDEL frequency looking across reads from all isolates, therefore further work needs to be done to explore the minor alleles in depth by lowering the thresholds and looking at unique diversity within isolate subgroups.

This study made comparative analyses of two genes, one CD8 T cell target antigen and one CD4 T cell target antigen, in *T. parva* parasites from cattle and buffalo in the SENAPA study area. Allelic diversity was investigated at the population level and more extensive diversity was observed in buffalo-derived parasites than cattle-maintained, for both antigen genes. Tp1 showed a high degree of allelic diversity, resulting in epitope variants. This variation is likely potential for immune evasion by escape from CD8 T cell recognition. Although there was greater allelic diversity observed in N60, there was no variance at the protein level, demonstrating it to be a conserved gene. An antigen gene that is conserved across the parasite population should be explored with potential as a vaccine candidate.

Chapter 5: General Discussion

The project aimed to assess epidemiology of *Theileria parva* around the Serengeti National Park (SENAPA), with a focus on understanding farmer perceptions of the disease and implementation of control methods, and the development and application of a long read sequencing approach to selected antigen genes in order to understand *T. parva* population genetics in cattle and buffalo. The key outcomes of the project were:

- The point prevalence of *T. parva* in the SENAPA area in 2016 was 5.07% (CI: 3.70-7.00%, n = 770) in cattle and 100% (CI: 85.00-100.00%, n = 22) in buffalo.
- The questionnaire survey established that farmers are concerned about ticks and as such, are using extensive amounts of acaricide, often incorrectly.
- A pipeline was successfully generated to amplify and analyse full length or near-full length sequences of target antigen genes, using long-read sequencing, demonstrating that:
    - Both antigen genes studied, Tp1 and N60, were polymorphic at the nucleotide level, but only Tp1 exhibited amino acid and therefore antigenic polymorphism.
    - There was greater diversity in buffalo-derived *T. parva* parasites than in cattle-maintained parasites in the SENAPA study area, and indications of population substructuring based upon Tp1.

5.1 Prevalence at the livestock / wildlife interface

The study established the point prevalence of *T. parva* to be 5.07% in cattle in 2016 in the SENAPA study area by means of cross-sectional sampling (n = 770). This is lower than has been reported in other hyperendemic areas in Tanzania, where the prevalence has been reported to be up to 50% (Kazungu et al. 2015a, Kazungu et al. 2015b, Kerario et al. 2017, Laisser et al. 2014) by nested p104 PCR. The Infectious Diseases of East African Livestock (IDEAL)

project (de Clare Bronsvoort et al. 2013) investigated the epidemiology of infectious cattle diseases in 548 cohort calves in western Kenya, and almost all calves were observed to be positive for *T. parva* throughout the 51 weeks of observation. In the study, 88 deaths were reported, of which 32 were confirmed due to ECF (36%). The study reported 70% of farmers using tick control. A study in Uganda by Muhanguzi et al. (2010a) established *T. parva* prevalence of 24% in cattle in varying age groups and management systems. The 100% prevalence of *T. parva* observed in the buffalo samples from the SENAPA study area was not unexpected, as buffalo in endemic areas are generally all found to be infected (Young et al. 1978). A study by Oura et al. (2011a) screened buffalo in four national parks in Uganda and revealed all buffalo to be *T. parva* positive in Lake Mburo National Park, near the border with Tanzania. In the two parks in northern Uganda (Murchiston Falls National Park and Kidepo National Park), however, all buffalo were negative for *T. parva.*

All of the cattle sampled in the survey were reported as being healthy, and displayed no signs of ECF at time of sampling. The *T. parva* positive cattle were therefore in all likelihood predominantly carrier animals. All of the cattle in the study area were native indigenous breeds, mostly Tarime zebu and Maswa zebu cross-breeds. It is thought that Tarime zebu cattle are an ECF-tolerant breed, acting as carriers of infection to more susceptible breeds present in mixed herds but not showing clinical signs of disease (Laisser et al. 2014, Ngowi 2008). The genetic potential for tolerance to ECF and other TBDs in Tarime cattle could be an important consideration in farmer selection of cattle breeds resilient to disease.

The low prevalence of infection observed could suggest a state of endemic stability, where cattle are regularly exposed to low levels of infection (not easily detected by PCR due to low parasitaemia) but have developed immunity and thus do not show signs of disease (Moll et al. 1984, Moll et al. 1986). Kivaria et al. (2004) carried out a study in Uganda to establish the endemic status of *T. parva* infections; in 96% of herds, antibody prevalence was above 70% in

six month old calves and calf mortality due to ECF was between 0% and 5.4%. Tick control practices were considered inefficient in the study area and it was determined that the study area was one of endemic stability. A few localised parts of the east African highlands are considered to have ECF endemic stability; extensive management conditions are used with very minimal acaricide use (Perry 1994).

Although farmers in the SENAPA study reported regularly seeing ticks, very few *R. appendiculatus* were observed on the cattle sampled. Although hard to confirm or quantify due to self-reporting, farmers reported cases of ECF and deaths due to ECF in their herds (mean cases 6.25, mean deaths 1.59, in the year preceding the questionnaire). It is, therefore, likely that without regular tick infestation, the cattle in the sampling area are not being regularly exposed to *T. parva*, and that the study area is one of endemic instability. This could be confirmed by looking at antibody status to quantify exposure.

5.2 Control at the livestock / wildlife interface

The questionnaire survey established that almost all (97.5%) of the farmers in the study area were using acaricides as the sole form of tick control and there was often very frequent use reported, with the most common interval of 5-7 days (37.89%) and as often as every 4 days in some cases (2.78%). It is, therefore, highly likely that the high levels of acaricide being used in the study area are responsible for the low tick counts observed on the sampled cattle and consequently the low prevalence of *T. parva* infection in cattle. However, exploratory tick sampling of vegetation within the Serengeti National Park, including in the central area (Seronera) that is remote from settled areas and cattle populations, also demonstrated very low tick counts. Although acaricide-treated cattle are known to graze in SENAPA, particularly in the areas close to the park boundary, acaricide could not be the cause of low tick counts in Seronera. A likely explanation could be seasonal pattern; tick sampling was attempted during the months of January and July which are typically both the beginning of the dry seasons, although farmers had reported this as a peak

time for ticks. Although the rainfall pattern is less distinct in Northern Tanzania than in the south, it is during the rainy season that tick activity and abundance would generally be expected to be highest (Walker 2003). It is likely, therefore, that the low tick counts observed within the protected area was a seasonal effect and sampling during other times of the year would likely result in significantly greater counts observed.

As well as the incorrect frequency of acaricide application, the survey also established that most of the farmers were underdosing their cattle. The use of and dependence upon acaricides can create a highly vulnerable ecological and epidemiological balance. Where such high levels of acaricide are being used, as often as every 4 days on some farms in this study, cattle can become highly susceptible to TBDs. Inappropriate use of acaricides, as well as the continual use of just one class of acaricide, is also highly likely to drive the development of resistance in ticks, resulting in consequences that could be highly detrimental to the herd and the livelihood of the farmer. There is, therefore, a real requirement for improved acaricide use by farmers. Manufacturers must improve the quality of the directions for use on their packaging, and extension officers have an important role in clarifying and emphasising to farmers the importance of correct acaricide use and the negative effects of inappropriate use, if the onset of tick resistance is to be prevented, or at least delayed. From the workshops, it was clear that although most farmers went to extension officers for advice, many villages did not have an extension officer and, therefore, it was likely that some farmers could not easily access advice. Significant associations were made between farmers of larger herds and the awareness of ECF, as well as farmers of larger herds being significantly more likely to spray their cattle. Levels of education were not assessed but it is likely that farmers of smaller herds may have less access to information and therefore are likely more vulnerable. Since privatisation of the veterinary services, control of endemic diseases such as ECF are the responsibility of the private sector. However, government policy makers have the responsibility, along with farmers, to provide and maintain dip tanks in order to help farmers control TBDs, as well as managing and advising on the correct

use of acaricides in order to reduce the development of resistance in ticks. Current government advice to farmers is for the use of pyrethroids in areas where both ticks and tsetse are present, and in areas where only ticks are present, non-pyrethroid products, such as Amitraz should be used (personal communication, Joyce Daffa, Tsetse Control Division). Indeed, in areas where ticks and tsetse occur, the government was incentivising the use of pyrethroids by subsidising the purchase price (personal communication, Joyce Daffa, Tsetse Control Division), which likely explains why all farmers who were spraying or dipping were using the cypermethrin products.

By the application of pyrethroids, or any effective acaricide, there is disruption of the balance between a vector and its host. If calves, in particular, are not exposed to infection, subsequent movement to areas with higher tick infestation, can result in lethal infection of *T. parva*, as well as a multitude of other TBDs (Seifert 1996). It therefore makes sense for calves to be exposed at an early age, when residing in an endemic area, and basic tick control methods can be used, such as hand-picking ticks, in order to prevent infestations but to allow exposure in order for the development of immunity.

Although the control options for ECF and other TBDs are limited to farmers in Tanzania, the study has highlighted significant scope for improved use of acaricides in the study area. It was observed that acaricide preparation is not often correct (least compliance was 0% farmers using Tantix correctly and best compliance was with Albadip, with 43.48% using correctly) which could be selecting for tick resistance faster than necessary. The practice of drug rotation could also be investigated; farmers are dependent on which products are available to them but policy makers could take action to minimise the development of resistance and, in combination with drug manufacturers, could be encouraging farmers to use more than a single acaricide drug class. It has been demonstrated that drug rotation can delay the development of resistance (Jonsson et al. 2010, Thullner et al. 2007). Additionally, farmers should consider a more seasonal use of acaricides rather than using a blanket treatment all year round, although in Northern Tanzania this can indeed be

difficult due to the less well defined seasonal rainfall patterns. In Tanzania it is a legal requirement to treat cattle with insecticide, with the guidelines of every 14 days during the dry season and every 7 days during the wet season (personal communication, Joyce Daffa, Tsetse Control Division), although this is not being adhered to in the study area.

When interviewed in 2016, none of the farmers in the study area reported using the ITM vaccine, with most being unaware of its existence. During the feedback workshops in 2017, it was apparent that some of the farmers were starting to hear reports of ITM vaccine use in other areas and there was some interest in the possibility of it being used, although concern was expressed about the cost. The workshops also demonstrated that the use of acaricide was farmer-led, despite there being government policies and subsidies for acaricide use. It is clear, therefore, that farmers discuss disease concerns and control options and can influence the decisions of their peers. In other areas of Tanzania there are reports of increased demand for the ITM vaccine (Di Giulio et al. 2009). It has been reported that use of the ITM vaccine reduced calf mortality rates from 80% to 2% in some pastoral areas of northern Tanzania (MacMillan 2014) and that higher prices can be achieved for cattle with the distinctive circular ear-tags identifying them to have been immunised against ECF. This increase in income has allowed pastoral farmers to diversify and to send more children to school, including girls (MacMillan 2014). The immunisation cost is between $6 and $14 per animal (personal communication, Livestock Officer Emmanuel Sindoya) (Di Giulio et al. 2009, Lynen 2005) and although full cost analyses have not been carried out yet for the study area, in other areas of Tanzania it has been reported that the ITM accounts for 1% of the total TBD costs, whereas acaricide application accounts for 14% of annual costs (Kivaria 2006).

The regional-specific nature of disease epidemiology is of great importance to consider and thus region-specific control strategies should be developed. There is wide variation in *T.parva* prevalence reported by other studies across Tanzania (Kazungu et al. 2015a, Kazungu et al. 2015b, Kerario et al. 2017,

Laisser et al. 2014) and a variety of farming systems that are operated, so there can be no 'one-size-fits-all' approach to control of vectors and TBDs. A particular strategy by governments and farmers must be based on both epidemiological and socioeconomic factors in that region. In order to establish the potential suitability of the ITM vaccine in different regions, there is a requirement for further understanding of how genetic and antigenic diversity in *T. parva* relates to vaccine protection. Extensive genotyping studies would ideally be carried out in order to establish the level of diversity in cattle (and buffalo if shared grazing) in combination with immunising cattle with the ITM vaccine and challenging to assess protection. Vaccine use must be justified based on farming system and livestock management, in conjunction with the epidemiology of the disease and the genetic 'suitability' of the *T. parva* field population, and as with any control method, must be cost-effective.

5.3 *T. parva* diversity at the livestock / wildlife interface

The study established and validated a genotyping pipeline to analyse genetic and antigenic diversity of *T. parva*. This is the first reported application of long-read sequencing to assess the diversity of *T. parva*, which allowed for sequencing of full-length or near full-length genes. Analysis of sequence polymorphisms of antigen genes Tp1 and N60 was carried out on cattle and buffalo samples from the SENAPA study area; the sampled buffalo inhabited the protected area of the National Park and cattle were farmed in villages within 5 km from the unfenced park boundary. Extensive genetic diversity of *T. parva* parasites was observed at the nucleotide level for both Tp1 and N60. In Tp1, 86 gene alleles were identified and in N60, 216 gene alleles were identified. The single epitope previously identified in Tp1 had four variants observed in the SENAPA *T. parva* population, which have all been previously reported (Elisa et al. 2015, Graham et al. 2008, Kerario et al. 2019, Pelle et al. 2011). In N60, however, the amino acid level was completely conserved. There is a strain-specific immune response of CD8 T cells in ITM-immunised cattle which is a result of antigen polymorphism (Morrison et al. 2015), therefore,

investigation of a conserved antigen as a vaccine candidate makes sense as protection could potentially be achieved against all *T. parva* parasite strains, both buffalo- and cattle-derived, which would be highly beneficial at a livestock/wildlife interface. The resolution of sequencing in this study was limited for N60 due to the low number of variable loci (14 nucleotides out of 983 bp) and its complete conservation at the amino acid level. However, the pipeline has been shown to be very useful for the analysis of polymorphic antigens, as demonstrated by Tp1.

The genetic diversity observed in the *T. parva* population was greater in buffalo-derived populations compared to the cattle-derived population, in agreement with previous studies (Bishop et al. 1994a, Kerario et al. 2019, Oura et al. 2004a, Oura et al. 2011b, Pelle et al. 2011). For both Tp1 and N60, most alleles were unique to either cattle or buffalo, with only relatively few shared (of 86 alleles in Tp1, 19 unique to cattle and 55 unique to buffalo, and of 216 alleles in N60,11 unique to cattle and 191 unique to buffalo). As well as at the population level, there was a high level of diversity within individuals, especially so in the buffalo, with many allelic variants present in individual animals, indicating the acquisition of infection from multiple tick feeds. Despite the lack of physical division between them, how cattle and buffalo interact in this study site is largely unstudied. Buffalo densities are usually low in the northern boundary areas of the Park (Metzger et al. 2010) and the sampled buffalo locations were at least 10 km from the northern boundary. Cattle are farmed along the northern boundary and are known, from discussion with farmers, to graze inside the Park at times, albeit illegally. These findings indicate that some cattle will move through buffalo habitats, although the frequency that cattle would be exposed to ticks from buffalo is unknown and individual herd exposure is also unknown. The key to dynamics of the *T. parva* population between buffalo and cattle is clearly the tick vector, and in areas where both hosts occur there are likely to be ticks feeding on both the cattle and buffalo and therefore shared transmission of a variety of TBDs, including *T. parva*. *R. appendiculatus* is known to feed on a range of species including eland, sable antelope and greater kudu (Norval et al. 1982), but little is known about their

relative preference for specific species, including cattle versus buffalo. Bloodmeal studies would, therefore, be required to investigate this; studies have been carried out to assess host preferences of tsetse based on bloodmeal identifications (Anderson et al. 2011, Auty et al. 2016, Clausen et al. 1998, Hoppenheit et al. 2010, Muturi et al. 2011), however, bloodmeal analysis has proved difficult in tick species (Leger et al. 2015). It is known, however, that high levels of acaricide are being used in the study area, and the number of ticks on cattle were extremely low. It could be hypothesised that as the number of ticks having fed on cattle is reduced by acaricide use, the proportion of ticks having fed on buffalo could increase, increasing the risk of infection with the more severe buffalo-derived *T. parva*, as seen in a study by Sitt et al. (2015) in Ol Pejeta, Kenya, where cattle were introduced to a site containing buffalo only and 17 of the 24 cattle developed severe clinical signs and died from confirmed buffalo-derived *T. parva*. As well as this risk, as mentioned previously, high acaricide use is also likely to be reducing initial exposure to *T. parva* and thus any significant development of immunity to infection. This all points to a state of high endemic instability.

To reduce the dependence on acaricides for TBDs, it is necessary to consider an integrated control strategy, whereby several control methods are used in combination. The application of the ITM vaccine must be considered in such a strategy. As the ITM vaccine is a live vaccine, of course there is the concern about introducing vaccine *T. parva* strains into the study area and the impact that onward transmission of these strains could have on unvaccinated cattle. De Deken (2007) reported an outbreak of ECF that was thought to be the consequence of importing immunised cattle into the Comoros Islands from Tanzania. Naïve ticks fed on the imported cattle and were presumed to have transmitted *T. parva* to susceptible cattle. Active surveillance would therefore be highly advisable in any area on the introduction of a novel vaccine, and the pipeline developed in this study could be applicable to such surveillance, as it has identified a low presence of sequences identical to TpM in the sampled population.

The sequencing data demonstrated that despite no farmers utilising the vaccine in the study area, alleles identical to that in the reference genome and vaccine stock *T. parva* Muguga (TpM) were present in a few individuals – 9 cattle and 1 buffalo for Tp1, and 5 cattle and 1 buffalo for N60. The TpM epitope variant of Tp1 was present in 36 (76.6%) individuals. The presence of the TpM epitope of Tp1 had previously been identified in isolates in Kenya and South Africa (Hemmink et al. 2018, Pelle et al. 2011). TpM is one of the three component stocks of the Muguga Cocktail ITM vaccine and has been found to have a high level of similarity in antigen gene sequence as one of the other components, Serengeti-transformed stock (Hemmink 2014, Hemmink et al. 2016). The predominance of unique alleles, in terms of being found in only cattle or buffalo, indicated that there may be some divergence in *T. parva* populations in cattle and buffalo in the study area, with only relatively few alleles shared by both. The shared alleles could indicate mixing of disparate populations or indeed the predominant separation could indicate that a diverse parasite population evolved in buffalos before the arrival of cattle in Africa and that the cattle-maintained *T. parva* parasite population is a subset of the ancient buffalo-derived population, whereby the smaller cattle parasite population adapted in order to achieve transmission between cattle. If there are indeed different populations in cattle and buffalo, it could suggest that the ITM vaccine may not be very effective in this area. In combination with this, the antigenic diversity observed in Tp1 could also suggest that protection would be limited, given the limited diversity in the vaccine. Although no significant substructuring was evident, TpM clustered within the cattle-derived alleles in both Tp1 and N60, demonstrating a similarity between this vaccine stock and the cattle-derived *T. parva* population in the study area. Although there could be buffalo-derived parasites present within cattle that are not being detected due to the lack of piroplasms, it may be that the level was so low as not to cause clinical disease (as the sampled cattle were healthy, with a few exceptions) which could indicate that in fact the ITM immunisation would be protective to cattle exposed to predominantly cattle-derived strains. As stated, however, the vaccine has limited diversity and there is a drive for alternative

vaccine candidates. A more diverse vaccine cocktail has been proposed, however a conserved antigen such as N60 could have the potential for offering protection across all parasite strains.

While the data presented provided suggestive indications of substructuring, it was relatively limited in terms of power. Despite the large number of cattle sampled, only a relatively low proportion (5.07%) were positive for *T. parva*, and only a proportion of these resulted in successful amplification of antigen genes (in all likelihood because the samples represented low parasitaemic carrier state), and only a small number of sympatric buffalo samples were available. In order to more fully understand the population structure, ideally a much larger sample set of both *T. parva* positive cattle and buffalo would be obtained, and more polymorphic antigens identified for application to the PacBio approach. A greater understanding of the pattern of such diversity across the geographic range of *T. parva*, in particular focusing on the remaining areas where cattle and buffalo interact, would be particularly enlightening in terms of informing how the overall population is structured, and potentially how much a role buffalo-derived *T. parva* plays in disease epidemiology across the parasite's range (and therefore how much of a priority understanding the vaccine cross-protection in this scenario actually is).

5.4 A more integrated approach

In order to alleviate the dependence on acaricides in the study area, a more sustainable and integrated control strategy could combine use of the ITM vaccine with correct acaricide use, as well as appropriate herd management such as reducing inter-herd contact where possible. It would be anticipated that acaricide use could be reduced if used in conjunction with the ITM vaccine. A reduction in the development of tick resistance was reported after the introduction of the ITM vaccine to other areas in Tanzania, based on the reduced frequency of acaricide application, with the added benefit of reduced costs to livestock keepers (Lynen 1999). Farmers and policy-makers face a real challenge in establishing integrated control strategies that are suitable to

particular areas and economically viable. Economic analyses have demonstrated that incorporating the ITM vaccine into integrated control in almost all production systems is important and the vaccine contributes significantly to disease control overall (Gachohi et al. 2012).

The goal with most tick-borne diseases is surely to establish and maintain a state of relative endemic stability, whereby the mammalian host can exist with the tick vector and disease, without high levels of morbidity or mortality. This has been achieved in certain regions in eastern and southern Africa, by combining immunisation with tolerant cattle, moderate acaricide use and thus moderate tick burdens (Minjauw 2003). But striking this balance is a real challenge and must be based on firm understanding of the particular epidemiological setting. It is hoped that the findings in this study can make a significant contribution to the understanding of the intricate biology at the livestock/wildlife interface at the study area.

5.5 Future work

Data was generated for antigen Tp4 as well as Tp1 and N60 in Chapter 4. Analysis of this antigen gene will be carried out in the immediate future, along with further in-depth sequence analysis of Tp1 and N60 data.

Assessment of spatiotemporal trends in *T. parva* prevalence and diversity, in conjunction with patterns of acaricide use, would have been valuable in order to further characterise the epidemiological state of the SENAPA study area and so future work would ideally extend this study to larger sample sizes and further study areas, over multiple time points. Serum samples were collected during the cattle survey and it would be of interest to determine seroprevalence of *T. parva* in the study area, to provide further epidemiological information by establishing *T. parva* exposure, as well as current detectable infection. Sero-positivity has been identified as a protective effect against ECF mortality; Thumbi et al. (2014) showed that there was an 88% reduction in risk hazard

for ECF mortality in *T. parva* sero-positive individuals compared to those sero-negative.

As mentioned, control strategies must not only be effective and appropriate to specific regions, but they must be cost-effective as this is a very 'real life' limitation to the small-scale agro-pastoral farmer. It would therefore, be of great use to carry out economic analyses, specifically looking at costs of acaricide use, vaccine use and the costs of integrated strategies.

There is a real requirement for epidemiological data on *T. parva* infection within ticks as well as in the mammalian hosts. Spatial tick abundance and distribution studies are required in the area to allow for parasite diversity studies in the tick vector as well as the mammalian hosts. Tick sampling would ideally be carried out throughout the year so as to avoid likely seasonal effects impacting tick sampling strategies. Distribution patterns and genotyping of parasites can be incorporated into mathematical models in order to analyse transmission patterns in different epidemiological settings. There are very few published studies on modelling of ECF (Vajana et al. 2018) and with some of the important parameters required for modelling captured in the data, this study site could be a good candidate. In addition, studies on tick resistance mechanisms are also indicated, given the likely, and potentially imminent, emergence of resistance in the study area.

Two projects have been established in part due to findings from the questionnaire survey. Although not analysed by me, the samples collected in the cross-sectional cattle survey have been screened for trypanosome prevalence, and the questionnaire survey captured data on tsetse and trypanosomiasis control in the study area. These data have indicated declining trypanosome prevalence at the livestock/wildlife interface and extensive trypanocide use, as well as acaricide use. The two new projects aim to 1) investigate the hypothesis that the widespread use of pyrethroids by farmers is achieving control of human trypanosomiasis, and 2) investigate mechanisms of trypanocide resistance in livestock.

5.6 Conclusions

The findings of this study revealed low point prevalence of *T. parva* in cattle in the study area. Within the parasite population sampled there was extensive genetic diversity, more so in buffalo-derived parasites than cattle-derived. Extensive use of acaricide is being used in the study area, as the sole control method. The ITM vaccine is not being used in the study area, although sequencing data revealed the presence of alleles identical to the *T. parva* Muguga vaccine strain present in several cattle and buffalo in the area, which could be due to the presence of immunised cattle in the area, or may be due to selective pressures maintaining the sequence. Tp1 showed antigenic diversity in its epitope variants, whereas N60 was completely conserved. N60 could be investigated for protective immunity and a potential vaccine candidate antigen.

These findings contribute to the understanding of the occurrence and genetic diversity of *T. parva* at this study area in Northern Tanzania and will be useful in informing policy makers in regards to developing control strategies. In turn, this should then aid farmers in the development of livestock management practices to sustainably control ECF at this interface.

References

Addah, W., Baah, J., Tia, S. and Okine, E. (2009) Knowledge and practices of smallholder farmers and herdsmen in the use of acaricides and gastrointestinal anthelminthes in Ghana. *Livestock Research for Rural Development, 21*(11).

Agresti, A. (2002) *Categorical Data Analysis,* Wiley & Sons.

Akoolo, L., Pelle, R., Saya, R., Awino, E., Nyanjui, J., Taracha, E. L., Kanyari, P., Mwangi, D. M. and Graham, S. P. (2008) Evaluation of the recognition of *Theileria parva* vaccine candidate antigens by cytotoxic T lymphocytes from Zebu cattle. *Vet Immunol Immunopathol, 121*(3-4), pp. 216-21.

Allsopp, B. A. and Allsopp, M. T. (1988) *Theileria parva*: genomic DNA studies reveal intra-specific sequence diversity. *Mol Biochem Parasitol, 28*(1), pp. 77-83.

Allsopp, B. A., Baylis, H. A., Allsopp, M. T., Cavalier-Smith, T., Bishop, R. P., Carrington, D. M., Sohanpal, B. and Spooner, P. (1993) Discrimination between six species of *Theileria* using oligonucleotide probes which detect small subunit ribosomal RNA sequences. *Parasitology, 107 (Pt 2)*, pp. 157-65.

Allsopp, M. T., Theron, J., Coetzee, M. L., Dunsterville, M. T. and Allsopp, B. A. (1999) The occurrence of *Theileria* and *Cowdria* parasites in African buffalo (*Syncerus caffer*) and their associated *Amblyomma hebraeum* ticks. *Onderstepoort J Vet Res, 66*(3), pp. 245-9.

Ambrose, N., Lloyd, D. and Maillard, J. C. (1999) Immune responses to *Dermatophilus congolensis* infections. *Parasitol Today, 15*(7), pp. 295-300.

Anderson, N. E., Mubanga, J., Fevre, E. M., Picozzi, K., Eisler, M. C., Thomas, R. and Welburn, S. C. (2011) Characterisation of the wildlife reservoir community for human and animal trypanosomiasis in the Luangwa Valley, Zambia. *PLoS Negl Trop Dis, 5*(6), pp. e1211.

Aurrecoechea, C., Brestelli, J., Brunk, B. P., Fischer, S., Gajria, B., Gao, X., Gingle, A., Grant, G., Harb, O. S., Heiges, M., Innamorato, F., Iodice, J., Kissinger, J. C., Kraemer, E. T., Li, W., Miller, J. A., Nayak, V., Pennington, C., Pinney, D. F., Roos, D. S., Ross, C., Srinivasamoorthy, G., Stoeckert, C. J., Jr., Thibodeau, R., Treatman, C. and Wang, H. (2010) EuPathDB: a portal to eukaryotic pathogen databases. *Nucleic Acids Res, 38*(Database issue), pp. D415-9.

Auty, H., Cleaveland, S., Malele, I., Masoy, J., Lembo, T., Bessell, P., Torr, S., Picozzi, K. and Welburn, S. C. (2016) Quantifying Heterogeneity in Host-Vector Contact: Tsetse (*Glossina swynnertoni* and *G. pallidipes*) Host Choice in Serengeti National Park, Tanzania. *PLoS One, 11*(10), pp. e0161291.

Baldwin, C. L., Black, S. J., Brown, W. C., Conrad, P. A., Goddeeris, B. M., Kinuthia, S. W., Lalor, P. A., MacHugh, N. D., Morrison, W. I., Morzaria, S. P. and et al. (1988) Bovine T cells, B cells, and null cells are transformed by the protozoan parasite *Theileria parva*. *Infect Immun, 56*(2), pp. 462-7.

Baldwin, C. L., Goddeeris, B. M. and Morrison, W. I. (1987) Bovine helper T-cell clones specific for lymphocytes infected with *Theileria parva* (Muguga). *Parasite Immunol, 9*(4), pp. 499-513.

Baldwin, C. L., Iams, K. P., Brown, W. C. and Grab, D. J. (1992) *Theileria parva*: CD4+ helper and cytotoxic T-cell clones react with a schizont-derived antigen associated with the surface of *Theileria parva*-infected lymphocytes. *Exp Parasitol, 75*(1), pp. 19-30.

Barnett, S. F. (1957) Theileriasis control. *Bulletin of Epizootic Diseases of Africa, 5*, pp. 343-357.

Barnett, S. F. and Brocklesby, D. W. (1966) Susceptibility of African Buffalo (Syncerus Caffer) to Infection with Theileria Parva (Theiler 1904). *British Veterinary Journal, 122*(9), pp. 379-&.

Bioconda, P. (2019) Available: https://github.com/PacificBiosciences/pbbioconda.

Bishop, R., Geysen, D., Spooner, P., Skilton, R., Nene, V., Dolan, T. and Morzaria, S. (2001) Molecular and immunological characterisation of *Theileria parva* stocks which are components of the 'Muguga cocktail' used for vaccination against East Coast fever in cattle. *Vet Parasitol, 94*(4), pp. 227-37.

Bishop, R., Geysen, P.D., Skilton, R., Odongo, D., Nene, V. (2002) *Genomic polymorphism, sexual recombination and molecular epidemiology of Theileria parva. In Theileria. ,* Kluwer: Academic Press.

Bishop, R., Musoke, A., Morzaria, S., Gardner, M. and Nene, V. (2004) Theileria: intracellular protozoan parasites of wild and domestic ruminants transmitted by ixodid ticks. *Parasitology, 129*, pp. S271-S283.

Bishop, R., Sohanpal, B., Kariuki, D. P., Young, A. S., Nene, V., Baylis, H., Allsopp, B. A., Spooner, P. R., Dolan, T. T. and Morzaria, S. P. (1992) Detection of a carrier state in *Theileria parva*-infected cattle by the polymerase chain reaction. *Parasitology, 104 (Pt 2)*, pp. 215-32.

Bishop, R. P., Hemmink, J.D., Morrison, W.I., Weir, W., Toye, P.G., Sitt, T., Spooner, P.R., Musoke, A.J., Skilton, R.A., Odongo, D.O. (2015) The African buffalo parasite *Theileria*. sp. (buffalo) can infect and immortalize cattle leukocytes and encodes divergent orthologues of *Theileria parva* antigen genes. *International Journal for Parasitology: Parasites and Wildlife, 4*, pp. 333-342.

Bishop, R. P., Sohanpal, B. K., Allsopp, B. A., Spooner, P. R., Dolan, T. T. and Morzaria, S. P. (1993) Detection of polymorphisms among *Theileria parva* stocks using repetitive, telomeric and ribosomal DNA probes and anti-schizont monoclonal antibodies. *Parasitology, 107 (Pt 1)*, pp. 19-31.

Bishop, R. P., Sohanpal, B. K., Morzaria, S. P., Dolan, T. T., Mwakima, F. N. and Young, A. S. (1994b) Discrimination between *Theileria-Parva* and *T-Taurotragi* in the Salivary-Glands of *Rhipicephalus-Appendiculatus* Ticks Using Oligonucleotides Homologous to Ribosomal-Rna Sequences. *Parasitology Research, 80*(3), pp. 259-261.

Bishop, R. P., Spooner, P. R., Kanhai, G. K., Kiarie, J., Latif, A. A., Hove, T., Masaka, S. and Dolan, T. T. (1994a) Molecular characterization of *Theileria* parasites: application to the epidemiology of theileriosis in Zimbabwe. *Parasitology, 109 (Pt 5)*, pp. 573-81.

Blouin, E. F. and Stoltsz, W. H. (1989) Comparative infection rates of *Theileria parva lawrencei* in salivary glands of *Rhipicephalus appendiculatus* and *Rhipicephalus zambeziensis*. *Onderstepoort J Vet Res, 56*(3), pp. 211-3.

Boone, R. B., Thirgood, S. J. and Hopcraft, J. G. (2006) Serengeti wildebeest migratory patterns modeled from rainfall and new vegetation growth. *Ecology, 87*(8), pp. 1987-94.

Bouslikhane, M. (2015) Cross border movements of animals and animal products and their relevance to the epidemiology of animal diseases in Africa. *Conference of the OIE Regional Commission*.

Brocklesby, D. W. (1961) Morbidity and mortality rates in East Coast Fever (*Theileria parva* infection) and their application to drug screening procedures. *British Veterinary Journal, 117*, pp. 529-532.

Bronsvoort, B. M., Nfon, C., Hamman, S. M., Tanya, V. N., Kitching, R. P. and Morgan, K. L. (2004) Risk factors for herdsman-reported foot-and-mouth disease in the Adamawa Province of Cameroon. *Prev Vet Med, 66*(1-4), pp. 127-39.

Brown, W. C., Lonsdale-Eccles, J. D., DeMartini, J. C. and Grab, D. J. (1990) Recognition of soluble *Theileria parva* antigen by bovine helper T cell clones: characterization and partial purification of the antigen. *J Immunol, 144*(1), pp. 271-7.

Brown, W. C., Shaw, M. K., Conrad, P. A. and Dolan, T. T. (1989a) *Theileria parva*: reappearance of schizonts in infected lymphoblastoid cells treated with parvaquone is dependent on interleukin 2-like growth factors. *Exp Parasitol, 68*(3), pp. 308-25.

Brown, W. C., Sugimoto, C., Conrad, P. A. and Grab, D. J. (1989b) Differential Response of Bovine T-Cell Lines to Membrane and Soluble-Antigens of *Theileria parva* Schizont-Infected Cells. *Parasite Immunol, 11*(6), pp. 567-583.

Burridge, M. J., Morzaria, S. P., Cunningham, M. P. and Brown, C. G. (1972) Duration of immunity to East Coast fever (*Theileria parva* infection of cattle). *Parasitology, 64*(3), pp. 511-5.

Busby, A. T., Ayllon, N., Kocan, K. M., Blouin, E. F., De La Fuente, G., Galindo, R. C., Villar, M. and De La Fuente, J. (2012) Expression of heat shock proteins and subolesin affects stress responses, *Anaplasma phagocytophilum* infection and questing behaviour in the tick, *Ixodes scapularis*. *Med Vet Entomol, 26*(1), pp. 92-102.

Buse, A. (1982) The Likelihood Ratio, Wald and Lagrange Multiplier Tests. *The American Statistician, 36*, pp. 153-157.

Bustin, S. A. (2010) Why the need for qPCR publication guidelines?--The case for MIQE. *Methods, 50*(4), pp. 217-26.

Cai, J. J. (2008) PGEToolbox: A Matlab toolbox for population genetics and evolution. *J Hered, 99*(4), pp. 438-40.

Campbell, K. and Hofer, H. (1995) *People and wildlife: Spatial dynamics and zones of interaction, In Serengeti II: Dynamics, Management and Conservation of an Ecosystem,* Chicago: University of Chicago Press.

Casey-Bryars, M., Reeve, R., Bastola, U., Knowles, N. J., Auty, H., Bachanek-Bankowska, K., Fowler, V. L., Fyumagwa, R., Kazwala, R., Kibona, T., King, A., King, D. P., Lankester, F., Ludi, A. B., Lugelo, A., Maree, F. F., Mshanga, D., Ndhlovu, G., Parekh, K., Paton, D. J., Perry, B., Wadsworth, J., Parida, S., Haydon, D. T., Marsh, T. L., Cleaveland, S. and Lembo, T. (2018) Waves of endemic foot-and-mouth disease in eastern Africa suggest feasibility of proactive vaccination approaches. *Nat Ecol Evol, 2*(9), pp. 1449-1457.

Catalano, D., Biasibetti, E., Lynen, G., Di Giulio, G., De Meneghi, D., Tomassone, L., Valenza, F. and Capucchio, M. T. (2015) "Ormilo disease" a disorder of zebu cattle in Tanzania: bovine cerebral theileriosis or new protozoan disease? *Trop Anim Health Prod, 47*(5), pp. 895-901.

Chaisi, M. E., Sibeko, K. P., Collins, N. E., Potgieter, F. T. and Oosthuizen, M. C. (2011) Identification of *Theileria parva* and *Theileria* sp. (buffalo) 18S rRNA gene sequence variants in the African Buffalo (*Syncerus caffer*) in southern Africa. *Vet Parasitol, 182*(2-4), pp. 150-62.

Chambers, J. M., Cleveland, W. S., Kleiner, B. and Tukey, P. A. (1983) *Graphical Methods for Data Analysis,* Chapman and Hall.

Chenyambuga, S. W., Waiswa, C., Saimo, M., Ngumi, P. and Gwakisa, P.S. (2010) Knowledge and perceptions of traditional livestock keepers on tick-borne diseases and sero-prevalence of *Theileria parva* around Lake Victoria Basin. *Livestock Research for Rural Development, 22*(7).

Choi, B. C. and Pak, A. W. (2005) A catalog of biases in questionnaires. *Prev Chronic Dis, 2*(1), pp. A13.

Clausen, P. H., Adeyemi, I., Bauer, B., Breloeer, M., Salchow, F. and Staak, C. (1998) Host preferences of tsetse (Diptera : Glossinidae) based on bloodmeal identifications. *Med Vet Entomol, 12*(2), pp. 169-180.

Clopper, C. J. and Pearson, E. S. (1934) The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika, 26*, pp. 404-413.

Coetzer, J. A. and Tustin, R. C. (2004) *Infectious Diseases of Livestock,* Second ed.*,* Oxford University Press.

Coetzer, J. A. W., Thomson, G. R. and Tustin, R. C. (1994) *Infectious diseases of livestock with special reference to Southern Africa,* Cape Town ; New York: Oxford University Press.

Collins, N. E. and Allsopp, B. A. (1999) *Theileria parva* ribosomal internal transcribed spacer sequences exhibit extensive polymorphism and mosaic evolution: application to the characterization of parasites from cattle and buffalo. *Parasitology, 118 (Pt 6)*, pp. 541-51.

Connelley, T. K., MacHugh, N. D., Pelle, R., Weir, W. and Morrison, W. I. (2011) Escape from CD8+ T cell response by natural variants of an immunodominant epitope from *Theileria parva* is predominantly due to loss of TCR recognition. *J Immunol, 187*(11), pp. 5910-20.

Conover, W. J. (1971) *Practical Nonparametric Statistics,* John Wiley & Sons.

Conrad, P. A., ole-MoiYoi, O. K., Baldwin, C. L., Dolan, T. T., O'Callaghan, C. J., Njamunggeh, R. E., Grootenhuis, J. G., Stagg, D. A., Leitch, B. L. and Young, A. S. (1989) Characterization of buffalo-derived theilerial parasites with monoclonal antibodies and DNA probes. *Parasitology, 98 Pt 2*, pp. 179-188.

Conrad, P. A., Stagg, D. A., Grootenhuis, J. G., Irvin, A. D., Newson, J., Njamunggeh, R. E., Rossiter, P. B. and Young, A. S. (1987) Isolation of *Theileria* parasites from African buffalo *(Syncerus caffer)* and characterization with anti-schizont monoclonal antibodies. *Parasitology, 94 (Pt 3)*, pp. 413-423.

Criado-Fornelio, A., Gonzalez-del-Rio, M. A., Buling-Sarana, A. and Barba-Carretero, J. C. (2004) The "expanding universe" of piroplasms. *Vet Parasitol, 119*(4), pp. 337-345.

Cross, P. C., Lloyd-Smith, J. O. and Getz, W. M. (2005) Disentangling association patterns in fission-fusion societies using African buffalo as an example. *Animal Behaviour, 69*, pp. 499-506.

Cvijovic, I., Good, B. H. and Desai, M. M. (2018) The Effect of Strong Purifying Selection on Genetic Diversity. *Genetics, 209*(4), pp. 1235-1278.

Daubenberger, C. A., Taracha, E. L., Gaidulis, L., Davis, W. C. and McKeever, D. J. (1999) Bovine gammadelta T-cell responses to the intracellular protozoan parasite *Theileria parva*. *Infect Immun, 67*(5), pp. 2241-9.

De Castro, J. J. (1997) Sustainable tick and tick-borne disease control in livestock improvement in developing countries. *Veterinary Parastology, 71*, pp. 77-97.

de Clare Bronsvoort, B. M., Thumbi, S. M., Poole, E. J., Kiara, H., Auguet, O. T., Handel, I. G., Jennings, A., Conradie, I., Mbole-Kariuki, M. N., Toye, P. G., Hanotte, O., Coetzer, J. A. and Woolhouse, M. E. (2013) Design and descriptive epidemiology of the Infectious Diseases of East African Livestock (IDEAL) project, a longitudinal calf cohort study in western Kenya. *Bmc Veterinary Research, 9*, pp. 171.

De Deken, R., Martin, V., Saido, A., Madder, M., Brandt, J. and Geysen, D. (2007) An outbreak of East Coast Fever on the Comoros: a consequence of the import of immunised cattle from Tanzania? *Vet Parasitol, 143*(3-4), pp. 245-53.

De Meneghi, D., Stachurski, F. and Adakal, H. (2016) Experiences in Tick Control by Acaricide in the Traditional Cattle Sector in Zambia and Burkina Faso: Possible Environmental and Public Health Implications. *Front Public Health, 4*, pp. 1-11.

De Vos, V. and Bengis, R. G. (1994) The disease status of African buffalo in South Africa, in W. Van Hoven, H. Ebedes & A. Conroy (eds.), Wildlife ranching: A celebration of diversity Proceedings of the 3rd International Wildlife Ranching Symposium, Centre for Wildlife Management, University of Pretoria, Pretoria, October, 1992.

De Vos, V., Bengis, R. G., Kriek, N. P. J., Michel, A., Keet, D. F., Raath, J. P. and Huchzermeyer, H. F. K. A. (2001) The epidemiology of tuberculosis in free-ranging African buffalo (*Syncerus caffer*) in the Kruger National Park, South Africa. *Onderstepoort Journal of Veterinary Research, 68*(2), pp. 119-130.

Di Giulio, G., Lynen, G., Medley, G., O'Callaghan, C. J., Kiara, H. K. and McDermott, J. J. (2003) Characterisation of theileriosis in pastoral production systems in NorthernTanzania. In: Proceedings of the 10th International symposium for Veterinary Epidemiology and Economics, Vina del Mar, Chile, 17–21 November.

Di Giulio, G., Lynen, G., Morzaria, S., Oura, C. and Bishop, R. (2009) Live immunization against East Coast fever--current status. *Trends Parasitol, 25*(2), pp. 85-92.

Dobbelaere, D. A. and Kuenzi, P. (2004) The strategies of the *Theileria* parasite: a new twist in host-pathogen interactions. *Curr Opin Immunol, 16*(4), pp. 524-30.

Dobbelaere, D. A., Roditi, I. J., Coquerelle, T. M., Kelke, C., Eichhorn, M. and Williams, R. O. (1991) Lymphocytes infected with *Theileria parva* require both cell-cell contact and growth factor to proliferate. *Eur J Immunol, 21*(1), pp. 89-95.

Dobbelaere, D. A., Shapiro, S. Z. and Webster, P. (1985a) Identification of a surface antigen on *Theileria parva* sporozoites by monoclonal antibody. *Proc Natl Acad Sci U S A, 82*(6), pp. 1771-5.

Dobbelaere, D. A., Webster, P., Leitch, B. L., Voigt, W. P. and Irvin, A. D. (1985b) *Theileria parva*: expression of a sporozoite surface coat antigen. *Exp Parasitol, 60*(1), pp. 90-100.

Dohoo, I., Martin, W., Stryhn, H. (2009) *Veterinary Epidemiologic Research,* Second ed.*,* Canada: VER Inc.

Dolan, T. T. (1986) Chemotherapy of East Coast fever: the long term weight changes, carrier state and disease manifestations of parvaquone treated cattle. *J Comp Pathol, 96*(2), pp. 137-46.

Dolan, T. T. (1999) Dogmas and misunderstandings in East Coast fever. *Trop Med Int Health, 4*(9), pp. A3-11.

Dolan, T. T., Young, A. S., Leitch, B. L. and Stagg, D. A. (1984) Chemotherapy of East Coast fever: parvaquone treatment of clinical disease induced by isolates of Theileria parva. *Vet Parasitol, 15*(2), pp. 103-16.

Donner, A. and Klar, N. (2000) Design and analysis of cluster randomisation trials in health research. *International Journal of Epidemiology, 30*(2), pp. 407-408.

Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., Peluso, P., Rank, D., Baybayan, P., Bettman, B., Bibillo, A., Bjornson, K., Chaudhuri, B., Christians, F., Cicero, R., Clark, S., Dalal, R., Dewinter, A., Dixon, J., Foquet, M., Gaertner, A., Hardenbol, P., Heiner, C., Hester, K., Holden, D., Kearns, G., Kong, X., Kuse, R., Lacroix, Y., Lin, S., Lundquist, P., Ma, C., Marks, P., Maxham, M., Murphy, D., Park, I., Pham, T., Phillips, M., Roy, J., Sebra, R., Shen, G., Sorenson, J., Tomaney, A., Travers, K., Trulson, M., Vieceli, J., Wegener, J., Wu, D., Yang, A., Zaccarin, D., Zhao, P., Zhong, F., Korlach, J. and Turner, S. (2009) Real-time DNA sequencing from single polymerase molecules. *Science, 323*(5910), pp. 133-8.

Elisa, M., Hasan, S. D., Moses, N., Elpidius, R., Skilton, R. and Gwakisa, P. (2015) Genetic and antigenic diversity of Theileria parva in cattle in Eastern and Southern zones of Tanzania. A study to support control of East Coast fever. *Parasitology, 142*(5), pp. 698-705.

Emery, D. L. (1981) Adoptive transfer of immunity to infection with *Theileria parva* (East Coast fever) between cattle twins. *Res Vet Sci, 30*(3), pp. 364-7.

Estes, A. B., Kuemmerle, T., Kushnir, H., Radeloff, V. C. and Shugart, H. H. (2012) Land-cover change and human population trends in the greater Serengeti ecosystem from 1984-2003. *Biological Conservation, 147*(1), pp. 255-263.

Eugui, E. M. and Emery, D. L. (1981) Genetically restricted cell-mediated cytotoxicity in cattle immune to *Theileria parva*. *Nature, 290*(5803), pp. 251-4.

FAO (1984) *Ticks and Tick-Borne Disease Control - A Practical Field Manual, Tick Control (Vol 1),* Rome: Food & Agriculture Organisation of the United Nations

FAO (2004) Module 1. Ticks: Acaricide resistance: Diagnosis, Management and Prevention. in: Food & Agriculture Organisation of the United Nations.

Fawcett, D., Musoke, A. and Voigt, W. (1984) Interaction of sporozoites of *Theileria parva* with bovine lymphocytes in vitro. I. Early events after invasion. *Tissue Cell, 16*(6), pp. 873-84.

Fawcett, D. W., Doxsey, S., Stagg, D. A. and Young, A. S. (1982) The entry of sporozoites of *Theileria parva* into bovine lymphocytes in vitro. Electron microscopic observations. *Eur J Cell Biol, 27*(1), pp. 10-21.

Fijorek, K. a. S., A. (2012) Separation-Resistant and Bias-Reduced Logistic Regression: STATISTICA Macro. *Journal of Statistical Software, 47*.

Freeman, A. R., Hoggart, C. J., Hanotte, O. and Bradley, D. G. (2006) Assessing the relative ages of admixture in the bovine hybrid zones of Africa and the Near East using X chromosome haplotype mosaicism. *Genetics, 173*(3), pp. 1503-10.

Gachohi, J., Skilton, R., Hansen, F., Ngumi, P. and Kitala, P. (2012) Epidemiology of East Coast fever (*Theileria parva* infection) in Kenya: past, present and the future. *Parasit Vectors, 5*, pp. 1-13.

Gardner, M. J., Bishop, R., Shah, T., de Villiers, E. P., Carlton, J. M., Hall, N., Ren, Q., Paulsen, I. T., Pain, A., Berriman, M., Wilson, R. J., Sato, S., Ralph, S. A., Mann, D. J., Xiong, Z., Shallom, S. J., Weidman, J., Jiang, L., Lynn, J., Weaver, B., Shoaibi, A., Domingo, A. R., Wasawo, D., Crabtree, J., Wortman, J. R., Haas, B., Angiuoli, S. V., Creasy, T. H., Lu, C., Suh, B., Silva, J. C., Utterback, T. R., Feldblyum, T. V., Pertea, M., Allen, J., Nierman, W. C., Taracha, E. L., Salzberg, S. L., White, O. R., Fitzhugh, H. A., Morzaria, S., Venter, J. C., Fraser, C. M. and Nene, V. (2005) Genome sequence of *Theileria parva*, a bovine pathogen that transforms lymphocytes. *Science, 309*(5731), pp. 134-7.

George, J. E., Pound, J. M. and Davey, R. B. (2004) Chemical control of ticks on cattle and the resistance of these parasites to acaricides. *Parasitology, 129 Suppl*, pp. S353-S366.

Geysen, D., Bazarusanga, T., Brandt, J. and Dolan, T. T. (2004) An unusual mosaic structure of the PIM gene of *Theileria parva* and its relationship to allelic diversity. *Mol Biochem Parasitol, 133*(2), pp. 163-73.

Geysen, D., Bishop, R., Skilton, R., Dolan, T. T. and Morzaria, S. (1999) Molecular epidemiology of *Theileria parva* in the field. *Trop Med Int Health, 4*(9), pp. A21-7.

Gifford-Gonzalez, D. and Hanotte, O. (2011) Domesticating Animals in Africa: Implications of Genetic and Archaeological Findings. *Journal of World Prehistory, 24*(1), pp. 1-23.

Gitau, G. K., McDermott, J. J., Katende, J. M., O'Callaghan, C. J., Brown, R. N. and Perry, B. D. (2000) Differences in the epidemiology of theileriosis on smallholder dairy farms in contrasting agro-ecological and grazing strata of highland Kenya. *Epidemiol Infect, 124*(2), pp. 325-35.

Gitau, G. K., Perry, B. D. and McDermott, J. J. (1999) The incidence, calf morbidity and mortality due to *Theileria parva* infections in smallholder dairy farms in Murang'a District, Kenya. *Prev Vet Med, 39*(1), pp. 65-79.

Githaka, N., Konnai, S., Bishop, R., Odongo, D., Lekolool, I., Kariuki, E., Gakuya, F., Kamau, L., Isezaki, M., Murata, S. and Ohashi, K. (2014) Identification and sequence characterization of novel *Theileria* genotypes from the waterbuck (*Kobus defassa*) in a *Theileria parva*-endemic area in Kenya. *Vet Parasitol, 202*(3-4), pp. 180-93.

Goddeeris, B. M., Katende, J. M., Irvin, A. D. and Chumo, R. S. (1982) Indirect fluorescent antibody test for experimental and epizootiological studies on East coast fever (*Theileria parva* infection in cattle). Evaluation of a cell culture schizont antigen fixed and stored in suspension. *Res Vet Sci, 33*(3), pp. 360-5.

Goddeeris, B. M. and Morrison, W. I. (1988) Techniques for the generation, cloning and characterisation of bovine cytotoxic T cells specific for the protozoan *Theileria parva*. *Journal of Tissue Culture Methods, 11*, pp. 101-10.

Goddeeris, B. M., Morrison, W. I. and Teale, A. J. (1986a) Generation of bovine cytotoxic cell lines, specific for cells infected with the protozoan parasite *Theileria parva* and restricted by products of the major histocompatibility complex. *Eur J Immunol, 16*(10), pp. 1243-9.

Goddeeris, B. M., Morrison, W. I., Teale, A. J., Bensaid, A. and Baldwin, C. L. (1986b) Bovine cytotoxic T-cell clones specific for cells infected with the protozoan parasite *Theileria parva*: parasite strain specificity and class I major histocompatibility complex restriction. *Proc Natl Acad Sci U S A, 83*(14), pp. 5238-42.

Goddeeris, B. M., Morrison, W. I., Toye, P. G. and Bishop, R. (1990) Strain specificity of bovine *Theileria parva*-specific cytotoxic T cells is determined by the phenotype of the restricting class I MHC. *Immunology, 69*(1), pp. 38-44.

Grab, D. J., Baldwin, C. L., Brown, W. C., Innes, E. A., Lonsdale-Eccles, J. D. and Verjee, Y. (1992) Immune CD4+ T cells specific for *Theileria parva*-infected lymphocytes recognize a 24-kilodalton protein. *Infect Immun, 60*(9), pp. 3892-6.

Graf, J. F., Gogolewski, R., Leach-Bing, N., Sabatini, G. A., Molento, M. B., Bordin, E. L. and Arantes, G. J. (2004) Tick control: an industry point of view. *Parasitology, 129*, pp. S427-S442.

Graham, S. P., Honda, Y., Pelle, R., Mwangi, D. M., Glew, E. J., de Villiers, E. P., Shah, T., Bishop, R., van der Bruggen, P., Nene, V. and Taracha, E. L. (2007) A novel strategy for the identification of antigens that are recognised by bovine MHC class I restricted cytotoxic T cells in a protozoan infection using reverse vaccinology. *Immunome Res, 3*.

Graham, S. P., Pelle, R., Honda, Y., Mwangi, D. M., Tonukari, N. J., Yamage, M., Glew, E. J., de Villiers, E. P., Shah, T., Bishop, R., Abuya, E., Awino, E., Gachanja, J., Luyai, A. E., Mbwika, F., Muthiani, A. M., Ndegwa, D. M., Njahira, M., Nyanjui, J. K., Onono, F. O., Osaso, J., Saya, R. M., Wildmann, C., Fraser, C. M., Maudlin, I., Gardner, M. J., Morzaria, S. P., Loosmore, S., Gilbert, S. C., Audonnet, J. C., van der Bruggen, P., Nene, V. and Taracha, E. L. (2006) *Theileria parva* candidate vaccine antigens recognized by immune bovine cytotoxic T lymphocytes. *Proc Natl Acad Sci U S A, 103*(9), pp. 3286-91.

Graham, S. P., Pelle, R., Yamage, M., Mwangi, D. M., Honda, Y., Mwakubambanya, R. S., de Villiers, E. P., Abuya, E., Awino, E., Gachanja, J., Mbwika, F., Muthiani, A. M., Muriuki, C., Nyanjui, J. K., Onono, F. O., Osaso, J., Riitho, V., Saya, R. M., Ellis, S. A., McKeever, D. J., MacHugh, N. D., Gilbert, S. C., Audonnet, J. C., Morrison, W. I., van der Bruggen, P. and Taracha, E. L. (2008) Characterization of the fine specificity of bovine CD8 T-cell responses to defined antigens from the protozoan parasite *Theileria parva*. *Infect Immun, 76*(2), pp. 685-94.

Graham, S. P., Saya, R., Awino, E., Ngugi, D., Nyanjui, J. K., Hecker, R., Taracha, E. L. and Nene, V. (2007a) Immunostimulatory CpG oligodeoxynucleotides enhance the induction of bovine CD4+ cytotoxic T-lymphocyte responses against the polymorphic immunodominant molecule of the protozoan parasite *Theileria parva*. *Vet Immunol Immunopathol, 115*(3-4), pp. 383-9.

Grootenhuis, J. G. (1989) The role of wildlife in the epidemiology of cattle theileriosis. in Dolan, T. T., (ed.) *Theileriosis in Eastern, Central and Southern Africa: Proceedings of a Workshop on East Coast Fever Immunization, Held in Lilongwe, Malawi 20-22 September 1988*: The International Laboratory for Research on Animal Diseases. pp. 129-136.

Grootenhuis, J. G., Leitch, B. L., Stagg, D. A., Dolan, T. T. and Young, A. S. (1987) Experimental induction of *Theileria parva lawrencei* carrier state in an African buffalo (*Syncerus caffer*). *Parasitology, 94 (Pt 3)*, pp. 425-31.

Grootenhuis, J. G., Young, A. S., Dolan, T. T. and Stagg, D. A. (1979) Characteristics of *Theileria* species (eland) infections in eland and cattle. *Res Vet Sci, 27*(1), pp. 59-68.

Gubbels, J. M., de Vos, A. P., van der Weide, M., Viseras, J., Schouls, L. M., de Vries, E. and Jongejan, F. (1999) Simultaneous detection of bovine *Theileria* and *Babesia* species by reverse line blot hybridization. *J Clin Microbiol, 37*(6), pp. 1782-9.

Haji, I. J., Malele, I. and Namangala, B. (2014) Occurrence of haemoparasites in cattle in Monduli district, northern Tanzania. *Onderstepoort J Vet Res, 81*(1).

Hayashida, K., Abe, T., Weir, W., Nakao, R., Ito, K., Kajino, K., Suzuki, Y., Jongejan, F., Geysen, D. and Sugimoto, C. (2013) Whole-genome sequencing of *Theileria parva* strains provides insight into parasite migration and diversification in the African continent. *DNA Res, 20*(3), pp. 209-20.

Hecker, K. H. and Roux, K. H. (1996) High and low annealing temperatures increase both specificity and yield in touchdown and stepdown PCR. *Biotechniques, 20*(3), pp. 478-85.

Hedger, R. S. (1972) Foot-and-mouth disease and the African buffalo (*Syncerus caffer*). *J Comp Pathol, 82*(1), pp. 19-28.

Hemmink, J. D. (2014) *Antigenic diversity in Theileria parva in vaccine stabilate and African buffalo*. Published PhD, University of Edinburgh.

Hemmink, J. D., Sitt, T., Pelle, R., de Klerk-Lorist, L. M., Shiels, B., Toye, P. G., Morrison, W. I. and Weir, W. (2018) Ancient diversity and geographical sub-structuring in African buffalo *Theileria parva* populations revealed through metagenetic analysis of antigen-encoding loci. *Int J Parasitol*.

Hemmink, J. D., Weir, W., MacHugh, N. D., Graham, S. P., Patel, E., Paxton, E., Shiels, B., Toye, P. G., Morrison, W. I. and Pelle, R. (2016) Limited genetic and antigenic diversity within parasite isolates used in a live vaccine against *Theileria parva*. *Int J Parasitol*, pp. 495-506.

Hezron, E. N., Adrian, M. and Robinson, H. M. (2012) Tick infestations in extensively grazed cattle and efficacy trial of high-cis cypermethrin pour-on preparation for control of ticks in Mvomero district in Tanzania. *Bmc Veterinary Research, 8*.

Homewood, K., Lambin, E. F., Coast, E., Kariuki, A., Kikula, I., Kivelia, J., Said, M., Serneels, S. and Thompson, M. (2001) Long-term changes in Serengeti-Mara wildebeest and land cover: pastoralism, population, or policies? *Proc Natl Acad Sci U S A, 98*(22), pp. 12544-9.

Homewood, K., Rodgers, W. A. and Arhem, K. (1987) Ecology of Pastoralism in Ngorongoro Conservation Area, Tanzania. *Journal of Agricultural Science, 108*, pp. 47-72.

Homewood, K., Trench, P., Randall, S., Lynen, G. and Bishop, B. (2006) Livestock health and socio-economic impacts of a veterinary intervention in Maasailand: Infection-and-treatment vaccine against East Coast fever. *Agricultural Systems, 89*(2-3), pp. 248-271.

Hoppenheit, A., Steuber, S., Bauer, B., Ouma, E. M., Diall, O., Zessin, K. H. and Clausen, P. H. (2010) Host preference of tsetse: an important tool to appraise the Nagana risk of cattle in the cotton zone of Mali. *Wien Klin Wochenschr, 122 Suppl 3*, pp. 81-6.

Huang, Y., Niu, B., Gao, Y., Fu, L. and Li, W. (2010) CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics, 26*(5), pp. 680-2.

Hughes, E. C., Kibona, T. K., de Glanville, W. A., Lankester, F., Davis, A., Carter, R. W., de Jong, R. M. F., Nyasebwa, O. M., Claxton, J. R., Cleaveland, S. and Allan, K. J. (2019) *Taenia multiceps* coenurosis in Tanzania: a major and under-recognised livestock disease problem in pastoral communities. *Veterinary Record, 184*(6).

Huson, D. H. and Bryant, D. (2006) Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol, 23*(2), pp. 254-67.

Hutchings, M. R. and Harris, S. (1997) Effects of farm management practices on cattle grazing behaviour and the potential for transmission of bovine tuberculosis from badgers to cattle. *Vet J, 153*(2), pp. 149-62.

Irvin, A. D. M., W.I. (1987) *Immunopathology, immunology and immunoprophylaxis of Theileria infections. In Immune Responses in parasitic infections: immunology, immunopathology, and immunoprophylaxis, Protozoa,* CRC Press, Boca Raton, Florida.

Jayaraman, S. (2018a) *SJ_RevSeq_22052018.R. The Roslin Institute, University of Edinburgh, siddharth.jayaraman@roslin.ed.ac.uk*.

Jayaraman, S. (2018b) *SJ_TPAntigen_primer_trim.m. The Roslin Institute, University of Edinburgh, siddharth.jayaraman@roslin.ed.ac.uk*

Jayaraman, S. (2019c) *Tp1_Antigen_Diversity_27Feb2019.m. The Roslin Institute, University of Edinburgh, siddharth.jayaraman@roslin.ed.ac.uk*

Jayaraman, S. (2019d) *N60_Antigen_Diversity_7Mar2019.m. The Roslin Institute, University of Edinburgh, siddharth.jayaraman@roslin.ed.ac.uk*

Jayaraman, S. (2019e) *Tp1_06112018_alleles_NuclDiversity.m. The Roslin Institute, University of Edinburgh, siddharth.jayaraman@roslin.ed.ac.uk*.

Jayaraman, S., Harris, C., Paxton, E., Donachie, A. M., Vaikkinen, H., McCulloch, R., Hall, J. P. J., Kenny, J., Lenzi, L., Hertz-Fowler, C., Cobbold, C., Reeve, R., Michoel, T. and Morrison, L. J. (2019) Application of long read sequencing to determine expressed antigen diversity in *Trypanosoma brucei* infections. *PLoS Negl Trop Dis, 13*(4), pp. e0007262.

Jepson, C., Asch, D. A., Hershey, J. C. and Ubel, P. A. (2005) In a mailed physician survey, questionnaire length had a threshold effect on response rate. *J Clin Epidemiol, 58*(1), pp. 103-5.

Jongejan, F. and Uilenberg, G. (2004) The global importance of ticks. *Parasitology, 129 Suppl*, pp. S3-14.

Jonsson, N. N., Miller, R. J., Kemp, D. H., Knowles, A., Ardila, A. E., Verrall, R. G. and Rothwell, J. T. (2010) Rotation of treatments between spinosad and amitraz for the control of *Rhipicephalus* (*Boophilus*) *microplus* populations with amitraz resistance. *Vet Parasitol, 169*(1-2), pp. 157-64.

Kaare, M. T., Picozzi, K., Mlengeya, T., Fevre, E. M., Mellau, L. S., Mtambo, M. M., Cleaveland, S. and Welburn, S. C. (2007) Sleeping sickness--a re-emerging disease in the Serengeti? *Travel Med Infect Dis, 5*(2), pp. 117-24.

Kariuki, D. P., Young, A. S., Morzaria, S. P., Lesan, A. C., Mining, S. K., Omwoyo, P., Wafula, J. L. and Molyneux, D. H. (1995) *Theileria parva* carrier state in naturally infected and artificially immunised cattle. *Trop Anim Health Prod, 27*(1), pp. 15-25.

Kariuki, E. K., Penzhorn, B. L. and Horak, I. G. (2012) Ticks (Acari: Ixodidae) infesting cattle and African buffaloes in the Tsavo conservation area, Kenya. *Onderstepoort J Vet Res, 79*(1), pp. E1-4.

Katale, B. Z., Mbugi, E. V., Karimuribo, E. D., Keyyu, J. D., Kendall, S., Kibiki, G. S., Godfrey-Faussett, P., Michel, A. L., Kazwala, R. R., van Helden, P. and Matee, M. I. (2013) Prevalence and risk factors for infection of bovine tuberculosis in indigenous cattle in the Serengeti ecosystem, Tanzania. *Bmc Veterinary Research, 9*, pp. 267.

Katende, J., Morzaria, S., Toye, P., Skilton, R., Nene, V., Nkonge, C. and Musoke, A. (1998) An enzyme-linked immunosorbent assay for detection of *Theileria parva* antibodies in cattle using a recombinant polymorphic immunodominant molecule. *Parasitology Research, 84*(5), pp. 408-16.

Katzer, F., Lizundia, R., Ngugi, D., Blake, D. and McKeever, D. (2011) Construction of a genetic map for Theileria parva: identification of hotspots of recombination. *Int J Parasitol, 41*(6), pp. 669-75.

Katzer, F., Ngugi, D., Oura, C., Bishop, R. P., Taracha, E. L., Walker, A. R. and McKeever, D. J. (2006) Extensive genotypic diversity in a recombining population of the apicomplexan parasite *Theileria parva*. *Infect Immun, 74*(10), pp. 5456-64.

Kazungu, Y. E., Mwega, E., Neselle, M. O., Sallu, R., Kimera, S. I. and Gwakisa, P. (2015a) Seroprevalence and carrier state of Theileria parva in cattle under two tick control regimes in small-holder farming systems of Tanzania. *Livestock Research for Rural Development, 27*(6).

Kazungu, Y. E., Mwega, E., Neselle, M. O., Sallu, R., Kimera, S. I. and Gwakisa, P. (2015b) Incremental effect of natural tick challenge on the infection and treatment method-induced immunity against *T. parva* in cattle under agro-pastoral systems in Northern Tanzania. *Ticks Tick Borne Dis, 6*(5), pp. 587-91.

Kerario, II, Chenyambuga, S. W., Mwega, E. D., Rukambile, E., Simulundu, E. and Simuunza, M. C. (2019) Diversity of two *Theileria parva* CD8+ antigens in cattle and buffalo-derived parasites in Tanzania. *Ticks Tick Borne Dis*.

Kerario, II, Simuunza, M., Laisser, E. L. K. and Chenyambuga, S. (2018) Exploring knowledge and management practices on ticks and tick-borne diseases among agro-pastoral communities in Southern Highlands, Tanzania. *Vet World, 11*(1), pp. 48-57.

Kerario, II, Simuunza, M. C., Chenyambuga, S. W., Koski, M., Hwang, S. G. and Muleya, W. (2017) Prevalence and risk factors associated with *Theileria parva* infection in cattle in three regions of Tanzania. *Trop Anim Health Prod*.

Kerario, I. I., Muleya, W., Chenyambuga, S., Koski, M., Hwang, S.G. and Martin, S. (2017) Abundance and distribution of Ixodid tick species infesting cattle reared under traditional farming systems in Tanzania. *African Journal of Agricultural Research, 12*, pp. 286-299.

Kimaro, E. G. a. C., O.C. (2013) Potential impact of climate change on livestock production and health in East Africa: A review. *Livestock Research for Rural Development, 25*(7), pp. 116.

Kivaria, F. M. (2006) Estimated direct economic costs associated with tick-borne diseases on cattle in Tanzania. *Trop Anim Health Prod, 38*(4), pp. 291-9.

Kivaria, F. M., Heuer, C., Jongejan, F., Okello-Onen, J., Rutagwenda, T., Unger, F. and Boehle, W. (2004) Endemic stability for *Theileria parva* infections in Ankole calves of the Ankole ranching scheme, Uganda. *Onderstepoort J Vet Res, 71*(3), pp. 189-95.

Kivaria, F. M., Kapaga, A. M., Mbassa, G. K., Mtui, P. F. and Wani, R. J. (2012) Epidemiological perspectives of ticks and tick-borne diseases in South Sudan: cross-sectional survey results. *Onderstepoort J Vet Res, 79*(1), pp. E1-E10.

Kock, R. A. (2005) What is this infamous "wildlife/livestock disease interface?" A review of current knowledge for the African continent. *Conservation and Development Interventions at the Wildlife/Livestock Interface*, pp. 1-13.

Komba, G. L., Mundogo, J.G.M. (1991) East Coast fever immunization in Tanzania Mainland. in Dolan, T. T., (ed.) *Ticks and Tick-borne Disease Control. Proceedings of a joint OAU, FAO and ILRAD workshop held in Kampala, Uganda 12-14 September*.

Koressaar, T., Lepamets, M., Kaplinski, L., Raime, K., Andreson, R. and Remm, M. (2018) Primer3_masker: integrating masking of template sequence with primer design software. *Bioinformatics, 34*(11), pp. 1937-1938.

Koressaar, T. and Remm, M. (2007) Enhancements and modifications of primer design program Primer3. *Bioinformatics, 23*(10), pp. 1289-1291.

Korlach, J., Chief Scientific Officer, Pacific Biosciences (2015) *Understanding Accuracy in SMRT Sequencing*, Available: https://www.pacb.com/wp-content/uploads/2015/09/Perspective_UnderstandingAccuracySMRTSequencing1.pdf [Accessed April 2019 2109].

Kumar, S., Stecher, G. and Tamura, K. (2016) MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol, 33*(7), pp. 1870-4.

Laisser, E. L., Chenyambuga, S. W., Karimuribo, E. D., Msalya, G., Kipanyula, M. J., Mwilawa, A. J., Mdegela, R. H. and Kusiluka, L. J. (2017) A review on prevalence, control measure, and tolerance of Tanzania Shorthorn Zebu cattle to East Coast fever in Tanzania. *Trop Anim Health Prod, 49*(4), pp. 813-822.

Laisser, E. L., Kipanyula, M. J., Msalya, G., Mdegela, R. H., Karimuribo, E. D., Mwilawa, A. J., Mwega, E. D., Kusiluka, L. and Chenyambuga, S. W. (2014) Tick burden and prevalence of *Theileria parva* infection in Tarime zebu cattle in the lake zone of Tanzania. *Trop Anim Health Prod, 46*(8), pp. 1391-6.

Lawrence, J. A. (1979) The differential diagnosis of the bovine Theilerias of Southern Africa. *J S Afr Vet Assoc, 50*(4), pp. 311-3.

Lawrence, J. A., Norval, R. A. and Uilenberg, G. (1983) *Rhipicephalus zambeziensis* as a vector of bovine Theileriae. *Trop Anim Health Prod, 15*(1), pp. 39-42.

Leger, E., Liu, X., Masseglia, S., Noel, V., Vourc'h, G., Bonnet, S. and McCoy, K. D. (2015) Reliability of molecular host-identification methods for ticks: an experimental in vitro study with *Ixodes ricinus*. *Parasit Vectors, 8*, pp. 433.

Lynen, G., Bakuname, C., Sanka, P. (1999) Ticks and Tick-borne Survey in the Northern Region of Tanzania. *Proceedings of the 17th Scientific Conference of the Tanzania Vet. Association, Arusha, Tanzania*, pp. 24-31.

Lynen, G., Di Giulio G., Homewood, K., Reid, R. and Mwilawa, A. (2005) Deployment of a live ECF vaccine in pastoral areas: Lessons learned from Tanzania. in *4th All Africa Conference on Animal Agriculture and 31ts Annual Meeting of the Tanzania Society for animal Production*,Arusha, Tanzania. pp. 193-201.

Lynen, G., Yrjo-Koskinen, A. E., Bakuname, C., Di Giulio, G., Mlinga, N., Khama, I., Hanks, J., Taylor, N. M., James, A. D., McKeever, D., Peters, A. R. and Rushton, J. (2012) East Coast fever immunisation field trial in crossbred dairy cattle in Hanang and Handeni districts in northern Tanzania. *Trop Anim Health Prod, 44*(3), pp. 567-72.

Lynen, G., Zeman, P., Bakuname, C., Di Giulio, G., Mtui, P., Sanka, P. and Jongejan, F. (2007) Cattle ticks of the genera *Rhipicephalus* and *Amblyomma* of economic importance in Tanzania: distribution assessed with GIS based on an extensive field survey. *Exp Appl Acarol, 43*(4), pp. 303-19.

Lynen, G., Zeman, P., Bakuname, C., Di Giulio, G., Mtui, P., Sanka, P. and Jongejan, F. (2008) Shifts in the distributional ranges of *Boophilus* ticks in Tanzania: evidence that a parapatric boundary between *Boophilus microplus* and *B. decoloratus* follows climate gradients. *Exp Appl Acarol, 44*(2), pp. 147-64.

MacHugh, N. D., Connelley, T., Graham, S. P., Pelle, R., Formisano, P., Taracha, E. L., Ellis, S. A., McKeever, D. J., Burrells, A. and Morrison, W. I. (2009) CD8+ T-cell responses to *Theileria parva* are preferentially directed to a single dominant antigen: Implications for parasite strain-specific immunity. *Eur J Immunol, 39*(9), pp. 2459-69.

MacMillan, S. (2014) *A short history of the 'live' vaccine protecting Africa's cattle against East Coast fever*, Available: https://ilvac.net/2014/08/19/a-short-history-of-the-live-vaccine-protecting-africas-cattle-against-east-coast-fever/ 2019].

Macpherson, C. N. L. (1995) The effect of transhumance on the epidemiology of animal diseases. *Prev Vet Med, 25*(2), pp. 213-224.

Magona, J. W., Greiner, M. and Mehlitz, D. (2000) Impact of tsetse control on the age-specific prevalence of trypanosomosis in village cattle in southeast Uganda. *Trop Anim Health Prod, 32*(2), pp. 87-98.

Mans, B. J., Pienaar, R. and Latif, A. A. (2015) A review of *Theileria* diagnostics and epidemiology. *Int J Parasitol Parasites Wildl, 4*(1), pp. 104-18.

Mans, B. J., Pienaar, R., Latif, A. A. and Potgieter, F. T. (2011) Diversity in the 18S SSU rRNA V4 hyper-variable region of *Theileria* spp. in Cape buffalo (*Syncerus caffer*) and cattle from southern Africa. *Parasitology, 138*(6), pp. 766-79.

Marcellino, W. L., Salih, D. A., Julla, II and El Hussein, A. R. (2012) Seroprevalence of East Coast fever in Central Equatoria State, South Sudan. *Vet Ital, 48*(4), pp. 379-85.

Martins, S. B., Di Giulio, G., Lynen, G., Peters, A. and Rushton, J. (2010) Assessing the impact of East Coast Fever immunisation by the infection and treatment method in Tanzanian pastoralist systems. *Prev Vet Med, 97*(3-4), pp. 175-82.

Mazeri, S. (2019) *distance_to_polygon_boundary.Rmd. The Roslin Institute, University of Edinburgh, smazeri@exseed.ed.ac.uk*

Mbassa G K, M. F. O. K., Melau L S B, Mlangwa J E D, Silayo R S, Bimbita E N, Hayghaimo A A and Mbiha E R (2009) A financing system for the control of tick-borne diseases in pastoral herds: The Kambala (Tanzania) model. . *Livestock Research for Rural Development, 21*(3).

McCulloch, B., Kalaye, W. J., Tungaraza, R., Suda, B. J. and Mbasha, E. M. (1968) A study of the life history of the tick *Rhipicephalus appendiculatus*--the main vector of East Coast fever--with reference to its behaviour under field conditions and with regard to its control in Sukumaland, Tanzania. *Bull Epizoot Dis Afr, 16*(4), pp. 477-500.

McKeever, D. J. (2006) *Theileria parva* and the bovine CTL response: down but not out? *Parasite Immunol, 28*(7), pp. 339-45.

McKeever, D. J. (2009) Bovine immunity - a driver for diversity in *Theileria* parasites? *Trends Parasitol, 25*(6), pp. 269-76.

McKeever, D. J. and Morrison, W. I. (1990) *Theileria parva*: the nature of the immune response and its significance for immunoprophylaxis. *Rev Sci Tech, 9*(2), pp. 405-21.

McKeever, D. J., Taracha, E. L., Innes, E. L., MacHugh, N. D., Awino, E., Goddeeris, B. M. and Morrison, W. I. (1994) Adoptive transfer of immunity to *Theileria parva* in the CD8+ fraction of responding efferent lymph. *Proc Natl Acad Sci U S A, 91*(5), pp. 1959-63.

McKeever, D. J., Taracha, E. L., Morrison, W. I., Musoke, A. J. and Morzaria, S. P. (1999) Protective immune mechanisms against *Theileria parva*: evolution of vaccine development strategies. *Parasitol Today, 15*(7), pp. 263-7.

Metzger, K. L., Sinclair, A. R. E., Hilborn, R., Hopcraft, J. G. C. and Mduma, S. A. R. (2010) Evaluating the protection of wildlife in parks: the case of African buffalo in Serengeti. *Biodiversity and Conservation, 19*(12), pp. 3431-3444.

Meyerhans, A., Vartanian, J. P. and Wain-Hobson, S. (1990) DNA recombination during PCR. *Nucleic Acids Res, 18*(7), pp. 1687-91.

Michel, A. L. and Bengis, R. G. (2012) The African buffalo: a villain for inter-species spread of infectious diseases in southern Africa. *Onderstepoort J Vet Res, 79*(2), pp. 453.

Michel, A. L., Bengis, R. G., Keet, D. F., Hofmeyr, M., Klerk, L. M., Cross, P. C., Jolles, A. E., Cooper, D., Whyte, I. J., Buss, P. and Godfroid, J. (2006) Wildlife tuberculosis in South African conservation areas: implications and challenges. *Vet Microbiol, 112*(2-4), pp. 91-100.

Minami, T., Spooner, P. R., Irvin, A. D., Ocama, J. G., Dobbelaere, D. A. and Fujinaga, T. (1983) Characterisation of stocks of *Theileria parva* by monoclonal antibody profiles. *Res Vet Sci, 35*(3), pp. 334-40.

Ministry of Agriculture (2012) *National Sample Census of Agriculture 2007/2008*, Available: http://harvestchoice.org/sites/default/files/downloads/publications/Tanzania_2007-8_Vol_5t.pdf/Tanzania_2007-8_Vol_5t.pdf Vt: Regional Report: Mara Region].

Minjauw, B. a. M., A. (2003) *Tick-borne diseases and poverty. The impact of ticks and tick-borne diseases on the livelihood of small-scale and marginal livestock owners in India and eastern and southern Africa*.

MLFD (2011) Ministry of Livestock and Fisheries Development. Livestock Sector Development Programme. The United Republic of Tanzania, Dar es Salaam, pp 1-2.

MLFD (2015) Ministry of Livestock and Fisheries Development. Tanzania Livestock Modernization Initiatives (TLMI) for the period 2015/2016-2020-2021. Statement by the President of United Republic of Tanzania. *Available at http://www.mifugouvuvi.go.tz*.

Mogotsi, K., Kgosikoma, O. E. and Lubinda, K. F. (2016) Wildlife-livestock interface, veterinary cordon fence damage, lack of protection zones, livestock theft and owner apathy: Complex socio-ecological dynamics in Foot and Mouth disease control in southern Africa. *Pastoralism-Research Policy and Practice, 6*.

Moll, G., Lohding, A. and Young, A. S. (1984) Epidemiology of Theileriases in the Trans-Mara Division, Kenya - Husbandry and Disease Background and Preliminary Investigations on Theileriases in Calves. *Prev Vet Med, 2*(6), pp. 801-831.

Moll, G., Lohding, A., Young, A. S. and Leitch, B. L. (1986) Epidemiology of theileriosis in calves in an endemic area of Kenya. *Vet Parasitol, 19*(3-4), pp. 255-73.

Morrison, W. I. (1996) Influence of host and parasite genotypes on immunological control of *Theileria* parasites. *Parasitology, 112 Suppl*, pp. S53-66.

Morrison, W. I. (2007) The biological and practical significance of antigenic variability in protective T cell responses against *Theileria parva*. *Vet Parasitol, 148*(1), pp. 21-30.

Morrison, W. I. (2015) The aetiology, pathogenesis and control of theileriosis in domestic animals. *Rev Sci Tech, 34*(2), pp. 599-611.

Morrison, W. I., Buscher, G., Murray, M., Emery, D. L., Masake, R. A., Cook, R. H. and Wells, P. W. (1981) *Theileria parva*: kinetics of infection in the lymphoid system of cattle. *Exp Parasitol, 52*(2), pp. 248-60.

Morrison, W. I., Connelley, T., Hemmink, J. D. and MacHugh, N. D. (2015) Understanding the basis of parasite strain-restricted immunity to *Theileria parva*. *Annu Rev Anim Biosci, 3*, pp. 397-418.

Morrison, W. I., Goddeeris, B. M., Brown, W. C., Baldwin, C. L. and Teale, A. J. (1989) *Theileria parva* in cattle: characterization of infected lymphocytes and the immune responses they provoke. *Vet Immunol Immunopathol, 20*(3), pp. 213-37.

Morrison, W. I., Goddeeris, B. M., Teale, A. J., Groocock, C. M., Kemp, S. J. and Stagg, D. A. (1987) Cytotoxic T-cells elicited in cattle challenged with *Theileria parva* (Muguga): evidence for restriction by class I MHC determinants and parasite strain specificity. *Parasite Immunol, 9*(5), pp. 563-78.

Morrison, W. I., MacHugh, N. D. and Lalor, P. A. (1996a) Pathogenicity of *Theileria parva* is influenced by the host cell type infected by the parasite. *Infect Immun, 64*(2), pp. 557-62.

Morrison, W. I. and McKeever, D. J. (1998) Immunology of infections with *Theileria parva* in cattle. *Chem Immunol, 70*, pp. 163-85.

Morrison, W. I. and McKeever, D. J. (2006) Current status of vaccine development against *Theileria* parasites. *Parasitology, 133 Suppl*, pp. S169-87.

Morzaria, S. and Williamson, S. (1999) Live vaccines for *Theileria parva*: deployment in Eastern, Central and Southern Africa. . *Proceedings of an FAO, OAU-IBAR and ILRI Workshop Held 12 March 1997 ILRI, Nairobi, Kenya*

Morzaria, S. P., Irvin, A. D., Voigt, W. P. and Taracha, E. L. (1987b) Effect of timing and intensity of challenge following immunization against East Coast fever. *Vet Parasitol, 26*(1-2), pp. 29-41.

Mugisha, A., McLeod, A., Percy, R. and Kyewalabye, E. (2005) Strategies, effectiveness and rationale of vector-borne disease control in the pastoralist system of south-western Uganda. *Trop Anim Health Prod, 37*(6), pp. 479-89.

Muhammed, S. I., Lauerman, L. H., Jr. and Johnson, L. W. (1975) Effect of humoral antibodies on the course of *Theileria parva* infection (East Coast fever) of cattle. *Am J Vet Res, 36*(4 Pt.1), pp. 399-402.

Muhanguzi, D., Ikwap, K., Picozzi, K., Waiswa C. (2010b) Molecular characterisation of analplasma and ehrlichia species in different cattle breeds and Age groups in Mbarara district (Western Uganda) *International Journal of Animal and Veterinary Advances, 2*(3), pp. 76-88.

Muhanguzi, D., Matovu, E. and Waiswa, C. (2010a) Prevalence and characterisation of *Theileria* and *Babesia* species in cattle under different husbandry systems in western Uganda. *International Journal of Animal and Veterinary Advances, 2*(2), pp. 51-58.

Muhanguzi, D., Picozzi, K., Hatendorf, J., Thrusfield, M., Welburn, S. C., Kabasa, J. D. and Waiswa, C. (2014a) Prevalence and spatial distribution of *Theileria parva* in cattle under crop-livestock farming systems in Tororo District, Eastern Uganda. *Parasit Vectors, 7*, pp. 91.

Mukhebi, A. W., Perry, B. D. and Kruska, R. (1992) Estimated Economics of Theileriosis Control in Africa. *Prev Vet Med, 12*(1-2), pp. 73-85.

Muleya, W., Namangala, B., Simuunza, M., Nakao, R., Inoue, N., Kimura, T., Ito, K., Sugimoto, C. and Sawa, H. (2012) Population genetic analysis and sub-structuring of *Theileria parva* in the northern and eastern parts of Zambia. *Parasit Vectors, 5*.

Muraguri, G. R., Kiara, H. K. and McHardy, N. (1999) Treatment of East Coast fever: a comparison of parvaquone and buparvaquone. *Vet Parasitol, 87*(1), pp. 25-37.

Murrell, P. (2005) *R Graphics,* Boca Raton: Chapman and Hall.

Musoke, A., Morzaria, S., Nkonge, C., Jones, E. and Nene, V. (1992) A recombinant sporozoite surface antigen of *Theileria parva* induces protection in cattle. *Proc Natl Acad Sci U S A, 89*(2), pp. 514-8.

Muturi, C. N., Ouma, J. O., Malele, II, Ngure, R. M., Rutto, J. J., Mithofer, K. M., Enyaru, J. and Masiga, D. K. (2011) Tracking the feeding patterns of tsetse flies (*Glossina genus*) by analysis of bloodmeals using mitochondrial cytochromes genes. *PLoS One, 6*(2), pp. e17284.

Muwanika, V., Kabi, F. and Masembe, C. (2016) Population genetic structure of *Theileria parva* field isolates from indigenous cattle populations of Uganda. *Ticks Tick Borne Dis, 7*(2), pp. 291-7.

Mwaseba, D. L. a. K., K.J. (2017) Knowledge, attitude, and practices about tsetse control among communities neighbouring Serengeti National Park, Tanzania. *Heliyon, 3*(6).

Nene, V. and Morrison, W. I. (2016) Approaches to vaccination against *Theileria parva* and *Theileria annulata*. *Parasite Immunol*.

Nene, V., Musoke, A., Gobright, E. and Morzaria, S. (1996) Conservation of the sporozoite p67 vaccine antigen in cattle-derived *Theileria parva* stocks with different cross-immunity profiles. *Infect Immun, 64*(6), pp. 2056-61.

Ngowi, E. E., Chenyambuga, S.W. and Gwakisa, P.S. (2008) Socioeconomic values and traditional management practices of Tarime zebu cattle in Tanzania. *Livestock Research for Rural Development, 20*.

Ngumi, P. N., Lesan, A. C., Williamson, S. M., Awich, J. R., Morzaria, S. P., Dolan, T. T., Shaw, M. K. and Young, A. S. (1994) Isolation and preliminary characterisation of a previously unidentified *Theileria* parasite of cattle in Kenya. *Res Vet Sci, 57*(1), pp. 1-9.

Nicolaiewsky, T. B., Richter, M. F., Lunge, V. R., Cunha, C. W., Delagostin, O., Ikuta, N., Fonseca, A. S., da Silva, S. S. and Ozaki, L. S. (2001) Detection of *Babesia equi* (Laveran, 1901) by nested polymerase chain reaction. *Vet Parasitol, 101*(1), pp. 9-21.

Norling, M., Bishop, R. P., Pelle, R., Qi, W., Henson, S., Drabek, E. F., Tretina, K., Odongo, D., Mwaura, S., Njoroge, T., Bongcam-Rudloff, E., Daubenberger, C. A. and Silva, J. C. (2015) The genomes of three stocks comprising the most widely utilized live sporozoite *Theileria parva* vaccine exhibit very different degrees and patterns of sequence divergence. *Bmc Genomics, 16*(1), pp. 729.

Norval, R. A., Walker, J. B. and Colborne, J. (1982) The ecology of *Rhipicephalus zambeziensis* and *Rhipicephalus appendiculatus* (Acarina, Ixodidae) with particular reference to Zimbabwe. *Onderstepoort J Vet Res, 49*(4), pp. 181-90.

Norval, R. A. I., Perry, B. D. and Young, A. S. (1992) *The Epidemiology of Theileriosis in Africa,* London: Academic Press.

Nthiwa, D., Alonso, S., Odongo, D., Kenya, E. and Bett, B. (2019) A participatory epidemiological study of major cattle diseases amongst Maasai pastoralists living in wildlife-livestock interfaces in Maasai Mara, Kenya. *Trop Anim Health Prod, 51*(5), pp. 1097-1103.

Odongo, D. O., Oura, C. A., Spooner, P. R., Kiara, H., Mburu, D., Hanotte, O. H. and Bishop, R. P. (2006) Linkage disequilibrium between alleles at highly polymorphic mini- and micro-satellite loci of *Theileria parva* isolated from cattle in three regions of Kenya. *Int J Parasitol, 36*(8), pp. 937-46.

Odongo, D. O., Sunter, J. D., Kiara, H. K., Skilton, R. A. and Bishop, R. P. (2010) A nested PCR assay exhibits enhanced sensitivity for detection of *Theileria parva* infections in bovine blood samples from carrier animals. *Parasitology Research, 106*(2), pp. 357-65.

Ogden, N. H., Swai, E., Beauchamp, G., Karimuribo, E., Fitzpatrick, J. L., Bryant, M. J., Kambarage, D. and French, N. P. (2005) Risk factors for tick attachment to smallholder dairy cattle in Tanzania. *Prev Vet Med, 67*(2-3), pp. 157-70.

Okiria, R., Okuna, N. M., Magona, J. W. and Mayende, J. S. (2002) Sustainability of tsetse control by subsequent treatment of 10% of a previously treated Ugandan cattle population with 1% w/v deltamethrin. *Trop Anim Health Prod, 34*(2), pp. 105-14.

Olwoch, J. M., Reyers, B., Engelbrecht, F. A. and Erasmus, B. F. N. (2008) Climate change and the tick-borne disease, *Theileriosis* (East Coast fever) in sub-Saharan Africa. *Journal of Arid Environments, 72*(2), pp. 108-120.

Otte, M. J. and Gumm, I. D. (1997) Intra-cluster correlation coefficients of 20 infections calculated from the results of cluster-sample surveys. *Prev Vet Med, 31*(1-2), pp. 147-50.

Oura, C. A., Asiimwe, B. B., Weir, W., Lubega, G. W. and Tait, A. (2005) Population genetic analysis and sub-structuring of *Theileria parva* in Uganda. *Mol Biochem Parasitol, 140*(2), pp. 229-39.

Oura, C. A., Bishop, R., Asiimwe, B. B., Spooner, P., Lubega, G. W. and Tait, A. (2007) *Theileria parva* live vaccination: parasite transmission, persistence and heterologous challenge in the field. *Parasitology, 134*(Pt 9), pp. 1205-1213.

Oura, C. A., Bishop, R., Wampande, E. M., Lubega, G. W. and Tait, A. (2004a) The persistence of component *Theileria parva* stocks in cattle immunized with the 'Muguga cocktail' live vaccine against East Coast fever in Uganda. *Parasitology, 129*(Pt 1), pp. 27-42.

Oura, C. A., Bishop, R. P., Wampande, E. M., Lubega, G. W. and Tait, A. (2004b) Application of a reverse line blot assay to the study of haemoparasites in cattle in Uganda. *Int J Parasitol, 34*(5), pp. 603-13.

Oura, C. A., Odongo, D. O., Lubega, G. W., Spooner, P. R., Tait, A. and Bishop, R. P. (2003) A panel of microsatellite and minisatellite markers for the characterisation of field isolates of *Theileria parva*. *Int J Parasitol, 33*(14), pp. 1641-53.

Oura, C. A., Tait, A., Asiimwe, B., Lubega, G. W. and Weir, W. (2011a) Haemoparasite prevalence and *Theileria parva* strain diversity in Cape buffalo (*Syncerus caffer*) in Uganda. *Vet Parasitol, 175*(3-4), pp. 212-9.

Oura, C. A., Tait, A., Asiimwe, B., Lubega, G. W. and Weir, W. (2011b) *Theileria parva* genetic diversity and haemoparasite prevalence in cattle and wildlife in and around Lake Mburo National Park in Uganda. *Parasitology Research, 108*(6), pp. 1365-74.

PacBio (2015-2019) *SMRT Analysis Software*, Available: https://www.pacb.com/products-and-services/analytical-software/smrt-analysis/.

Patel, E. H., Lubembe, D. M., Gachanja, J., Mwaura, S., Spooner, P. and Toye, P. (2011) Molecular characterization of live *Theileria parva* sporozoite vaccine stabilates reveals extensive genotypic diversity. *Vet Parasitol, 179*(1-3), pp. 62-8.

Pearson, T. W., Lundin, L. B., Dolan, T. T. and Stagg, D. A. (1979) Cell-mediated immunity to *Theileria*-transformed cell lines. *Nature, 281*(5733), pp. 678-80.

Peeling, R. W., Smith, P. G. and Bossuyt, P. M. (2010) A guide for diagnostic evaluations. *Nat Rev Microbiol, 8*(12 Suppl), pp. S2-6.

Pegram, R. G. and Oosterwijk, G. P. (1990) The effect of *Amblyomma variegatum* on liveweight gain of cattle in Zambia. *Med Vet Entomol, 4*(3), pp. 327-30.

Pelle, R., Graham, S. P., Njahira, M. N., Osaso, J., Saya, R. M., Odongo, D. O., Toye, P. G., Spooner, P. R., Musoke, A. J., Mwangi, D. M., Taracha, E. L., Morrison, W. I., Weir, W., Silva, J. C. and Bishop, R. P. (2011) Two *Theileria parva* CD8 T cell antigen genes are more variable in buffalo than cattle parasites, but differ in pattern of sequence diversity. *PLoS One, 6*(4), pp. e19015.

Perry, B. D. (1994) Modelling vector-borne disease. Epidemiology and the impact of control programs, in Modeling vector-borne and other parasitic diseases. *Proceedings of a Workshop organised by ILRAD in Collaboration with FAO, ILRAD, Nairobi, Kenya 23-27 November 1992*.

Petney, T. N. and Andrews, R. H. (1998) Multiparasite communities in animals and humans: frequency, structure and pathogenic significance. *Int J Parasitol, 28*(3), pp. 377-93.

Pienaar, R., Latif, A. A., Thekisoe, O. M. and Mans, B. J. (2014) Geographic distribution of *Theileria* sp. (buffalo) and *Theileria* sp. (bougasvlei) in Cape buffalo *(Syncerus caffer)* in southern Africa: implications for speciation. *Parasitology, 141*(3), pp. 411-24.

Pienaar, R., Potgieter, F. T., Latif, A. A., Thekisoe, O. M. and Mans, B. J. (2011) Mixed *Theileria* infections in free-ranging buffalo herds: implications for diagnosing *Theileria parva* infections in Cape buffalo (*Syncerus caffer*). *Parasitology, 138*(7), pp. 884-95.

Potgieter, F. T., Stoltsz, W. H., Blouin, E. F. and Roos, J. A. (1988) Corridor disease in South Africa: a review of the current status. *J S Afr Vet Assoc, 59*(3), pp. 155-60.

R Core Team (2018) *R: A language and environment for statistical computing.*, Available: http://www.R-project.org/.

Radley, D. E. (1981) Infection and treatment method of immunisation against theileriosis. In: Irvin, A.D., Cunningham, M.P., Young, A.S. (Eds.). *Advances in the Control of Theileriosis: Proceedings of an International Conference Held at the International Laboratory for Research on Animal Diseases, Nairobi, 9-13 February 1981. MartinusNijhoff, The Hague*, pp. 227-237.

Radley, D. E., Brown, C.G.D., Burridge, M.J., Cunningham, M.P., Kirimi, I.M., Purnell, R.E. and Young, A.S. (1975a) East Coast Fever: 1. Chemoprophylactic Immunisation of cattle against *Theileria parva* (Muguga) and five theilerial strains. *Vet Parasitol, 1*, pp. 35-41.

Radley, D. E., Brown, C.G.D., Cunningham, M.P., Kimber, C.D., Musisi, F.L., Payne, R.C., Purnell, R.E., Stagg, S.M. and Young, A.S. (1975c) East Coast Fever: 3. Chemoprophylactic immunization of cattle using oxytetracycline and a combination of theilerial strains. *Vet Parasitol, 1*, pp. 51-60.

Radley, D. E., Young, A. S., Grootenhuis, J. G., Cunningham, M. P., Dolan, T. T. and Morzaria, S. P. (1979) Further-Studies on the Immunization of Cattle against *Theileria-Lawrencei* by Infection and Chemoprophylaxis. *Vet Parasitol, 5*(2-3), pp. 117-128.

Radley, D. E., Young, A.S., Brown, C.G.D., Burridge, M.J., Cunningham, M.P., Musisi, F.L. and Purnell, R.E. (1975b) East Coast Fever: 2. Cross-immunity trials with a Kenya strain of *Theileria Lawrencei*. *Vet Parasitol, 1*, pp. 43-50.

Rahman, M. S. and Sultana, M. (2017) Performance of Firth-and log*F*-type penalized methods in risk prediction for small or sparse binary data. *BMC Med Res Methodol, 17*(1), pp. 33.

Rhoads, A. and Au, K. F. (2015) PacBio Sequencing and Its Applications. *Genomics Proteomics Bioinformatics, 13*(5), pp. 278-89.

Rhyan, J. C. and Spraker, T. R. (2010) Emergence of Diseases From Wildlife Reservoirs. *Veterinary Pathology, 47*(1), pp. 34-39.

Riddles, P. W. and Nolan, J. (1987) Prospects for the management of arthropod resistance to pesticides. *Int J Parasitol, 17*(2), pp. 679-87.

Rubaire-Akiiki, C. M., Okello-Onen, J., Musunga, D., Kabagambe, E. K., Vaarst, M., Okello, D., Opolot, C., Bisagaya, A., Okori, C., Bisagati, C., Ongyera, S. and Mwayi, M. T. (2006) Effect of agro-ecological zone and grazing system on incidence of East Coast Fever in calves in Mbale and Sironko Districts of Eastern Uganda. *Prev Vet Med, 75*(3-4), pp. 251-66.

Sajid, M. S., Iqbal, Z., Khan, M. N., Muhamamd, G. and Iqbal, M. U. (2007) Effect of Hyalomma Ticks (*Acari: Ixodidae*) on Milk Production of Dairy Buffaloes *(Bos bubalus bubalis*) of Punjab (Pakistan). *Italian Journal of Animal Science, 6*, pp. 939-941.

Schadt, E. E., Turner, S. and Kasarskis, A. (2010) A window into third-generation sequencing. *Hum Mol Genet, 19*(R2), pp. R227-40.

Schloss, P. D. and Westcott, S. L. (2011) Assessing and improving methods used in operational taxonomic unit-based approaches for 16S rRNA gene sequence analysis. *Appl Environ Microbiol, 77*(10), pp. 3219-26.

Schmieder, R. and Edwards, R. (2011) Quality control and preprocessing of metagenomic datasets. *Bioinformatics, 27*(6), pp. 863-4.

Schnittger, L., Yin, H., Qi, B., Gubbels, M. J., Beyer, D., Niemann, S., Jongejan, F. and Ahmed, J. S. (2004) Simultaneous detection and differentiation of *Theileria* and *Babesia* parasites infecting small ruminants by reverse line blotting. *Parasitology Research, 92*(3), pp. 189-96.

Schoonman, L. and Swai, E. S. (2010) Herd- and animal-level risk factors for bovine leptospirosis in Tanga region of Tanzania. *Trop Anim Health Prod, 42*(7), pp. 1565-72.

Schreuder, B. E., Uilenberg, G. and Tondeur, W. (1977) Studies on Theileriidae (Sporozoa) in Tanzania. VIII. Experiments with African buffalo (Syncerus caffer). *Tropenmed Parasitol, 28*(3), pp. 367-71.

Seifert, H. S. H. (1996) *Tropical Animal Health,* Dordrecht, The Netherlands: Kluwer Academic Publishers.

Shakoor, O., Taylor, R. B. and Behrens, R. H. (1997) Assessment of the incidence of substandard drugs in developing countries. *Trop Med Int Health, 2*(9), pp. 839-45.

Shen, W., Le, S., Li, Y. and Hu, F. (2016) SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. *PLoS One, 11*(10), pp. e0163962.

Sibeko, K. P., Geysen, D., Oosthuizen, M. C., Matthee, C. A., Troskie, M., Potgieter, F. T., Coetzer, J. A. and Collins, N. E. (2010) Four p67 alleles identified in South African *Theileria parva* field samples. *Vet Parasitol, 167*(2-4), pp. 244-54.

Sibeko, K. P., Oosthuizen, M. C., Collins, N. E., Geysen, D., Rambritch, N. E., Latif, A. A., Groeneveld, H. T., Potgieter, F. T. and Coetzer, J. A. (2008) Development and evaluation of a real-time polymerase chain reaction test for the detection of *Theileria parva* infections in Cape buffalo (*Syncerus caffer*) and cattle. *Vet Parasitol, 155*(1-2), pp. 37-48.

Simuunza, M., Weir, W., Courcier, E., Tait, A. and Shiels, B. (2011) Epidemiological analysis of tick-borne diseases in Zambia. *Vet Parasitol, 175*(3-4), pp. 331-42.

Sinclair, A. R. E. a. A., P. (1995) *Serengeti II Dynamics, Management, and Conservation of an Ecosystem,* The University of Chicago.

Sitt, T., Henson, S., Morrison, W. I. and Toye, P. (2019) Similar levels of diversity in the gene encoding the p67 sporozoite surface protein of *Theileria parva* are observed in blood samples from buffalo and cattle naturally infected from buffalo. *Vet Parasitol, 269*, pp. 21-27.

Sitt, T., Poole, E. J., Ndambuki, G., Mwaura, S., Njoroge, T., Omondi, G. P., Mutinda, M., Mathenge, J., Prettejohn, G., Morrison, W. I. and Toye, P. (2015) Exposure of vaccinated and naive cattle to natural challenge from buffalo-derived *Theileria parva*. *Int J Parasitol Parasites Wildl, 4*(2), pp. 244-51.

Sivakumar, T., Hayashida, K., Sugimoto, C. and Yokoyama, N. (2014) Evolution and genetic diversity of *Theileria*. *Infect Genet Evol, 27*, pp. 250-63.

Skilton, R. A., Bishop, R. P., Katende, J. M., Mwaura, S. and Morzaria, S. P. (2002) The persistence of *Theileria parva* infection in cattle immunized using two stocks which differ in their ability to induce a carrier state: analysis using a novel blood spot PCR assay. *Parasitology, 124*(Pt 3), pp. 265-76.

Smeeth, L., Fletcher, A. E., Stirling, S., Nunes, M., Breeze, E., Ng, E., Bulpitt, C. J. and Jones, D. (2001) Randomised comparison of three methods of administering a screening questionnaire to elderly people: findings from the MRC trial of the assessment and management of older people in the community. *BMJ, 323*(7326), pp. 1403-7.

Smyth, R. P., Schlub, T. E., Grimm, A., Venturi, V., Chopra, A., Mallal, S., Davenport, M. P. and Mak, J. (2010) Reducing chimera formation during PCR amplification to ensure accurate genotyping. *Gene, 469*(1-2), pp. 45-51.

Snodgrass, D. R., Trees, A. J., Bowyer, W. A., Bergman, J. R., Daft, J. and Wall, A. E. (1972) East Coast Fever: field challenge of cattle immunised against *Theileria parva* (Muguga). *Trop Anim Health Prod, 4*(3), pp. 142-51.

Stagg, D. A., Bishop, R. P., Morzaria, S. P., Shaw, M. K., Wesonga, D., Orinda, G. O., Grootenhuis, J. G., Molyneux, D. H. and Young, A. S. (1994) Characterization of *Theileria parva* Which Infects Waterbuck (*Kobus defassa*). *Parasitology, 108*, pp. 543-554.

Stagg, D. A., Young, A. S., Leitch, B. L., Grootenhuis, J. G. and Dolan, T. T. (1983) Infection of mammalian cells with *Theileria* species. *Parasitology, 86 (Pt 2)*, pp. 243-54.

Sugimoto, C. and Fujisaki, K. (2002) *Non-transforming Theileria parasites of ruminants., Theileria,* Boston: Kluwer Academic Publishers.

Swai, E. S., French, N. P., Karimuribo, E. D., Fitzpatrick, J. L., Bryant, M. J., Brown, P. E. and Ogden, N. H. (2005) Spatial and management factors associated with exposure of smallholder dairy cattle in Tanzania to tick-borne pathogens. *Int J Parasitol, 35*(10), pp. 1085-96.

Swai, E. S., Karimuribo, E. D., Kambarage, D. M. and Moshy, W. E. (2009) A longitudinal study on morbidity and mortality in youngstock smallholder dairy cattle with special reference to tick borne infections in Tanga region, Tanzania. *Vet Parasitol, 160*(1-2), pp. 34-42.

Swai, E. S., Karimuribo, E. D., Kambarage, D. M., Moshy, W. E. and Mbise, A. N. (2007) A comparison of seroprevalence and risk factors for *Theileria parva* and *T. mutans* in smallholder dairy cattle in the Tanga and Iringa regions of Tanzania. *Vet J, 174*(2), pp. 390-6.

Swai, E. S., Mbise, A.N., Kessy, V., Kaaya, E., Sanka, P. and Loomu, P.M. (2005a) Farm constraints, cattle disease perception and tick management practices in pastoral Maasai community-Ngorongoro, Tanzania. *Livestock Research for Rural Development, 17*(2).

Taracha, E. L., Awino, E. and McKeever, D. J. (1997) Distinct CD4+ T cell helper requirements in *Theileria parva*-immune and -naive bovine CTL precursors. *J Immunol, 159*(9), pp. 4539-45.

Taracha, E. L., Goddeeris, B. M., Morzaria, S. P. and Morrison, W. I. (1995a) Parasite strain specificity of precursor cytotoxic T cells in individual animals correlates with cross-protection in cattle challenged with *Theileria parva*. *Infect Immun, 63*(4), pp. 1258-62.

Taracha, E. L., Goddeeris, B. M., Teale, A. J., Kemp, S. J. and Morrison, W. I. (1995b) Parasite strain specificity of bovine cytotoxic T cell responses to *Theileria parva* is determined primarily by immunodominance. *J Immunol, 155*(10), pp. 4854-60.

Taracha, E. L., Morrison, W. I. and Goddeeris, B. M. (1992) The cytotoxic T cell: MHC restriction, strain specificity and role in immunity to *Theileria parva* infection. *Ann Rech Vet, 23*(3), pp. 311-8.

Tarimo, M. A. (2013) *Studies on the Prevalence of East Coast Fever among cattle in Kilosa District*. Published Master of Science, Sokoine University of Agriculture, Morogoro, Tanzania.

Tatchell, R. J. a. E., E. (1986) Tick (Acari: Ixodidae) ecological studies in Tanzania. *Bulletin of Entomological Research, 76*, pp. 229-246.

TAWIRI (2014) *Aerial Total Count of Elephant and Buffalo in the Serengeti Ecosystem. Tanzania Wildlife Research Institute in collaboration with Tanzania National Parks and Frankfurt Zoological Society.*

TAWIRI (2016) *Tanzania Wildlife Research Institute, 2016, Wildlife, Livestock and Bomas census in the Serengeti Ecosystem, Dry Season, 2016. TAWIRI Aerial Survey Report.*

Tebaldi, G., Williams, L. B., Verna, A. E., Macchi, F., Franceschi, V., Fry, L. M., Knowles, D. P. and Donofrio, G. (2017) Assessment and optimization of *Theileria parva* sporozoite full-length p67 antigen expression in mammalian cells. *PLoS Negl Trop Dis, 11*(8), pp. e0005803.

Thomson, G. R. (1995) Overview of foot and mouth disease in southern Africa. *Rev Sci Tech, 14*(3), pp. 503-20.

Thrusfield, M. (2005) *Veterinary Epidemiology,* Third ed.*,* Blackwell Science Ltd.

Thullner, F., Willadsen, P. and Kemp, D. (2007) Acaricide rotation strategy for managing resistance in the tick *Rhipicephalus* (*Boophilus*) microplus (Acarina: Ixodidae): laboratory experiment with a field strain from Costa Rica. *J Med Entomol, 44*(5), pp. 817-21.

Thumbi, S. M., Bronsvoort, B. M., Poole, E. J., Kiara, H., Toye, P. G., Mbole-Kariuki, M. N., Conradie, I., Jennings, A., Handel, I. G., Coetzer, J. A., Steyl, J. C., Hanotte, O. and Woolhouse, M. E. (2014) Parasite co-infections and their impact on survival of indigenous cattle. *PLoS One, 9*(2), pp. e76324.

Toye, P. G., Metzelaar, M. J., Wijngaard, P. L. J., Nene, V., Iams, K., Roose, J., Nyanjui, J. K., Gobright, E., Musoke, A. J. and Clevers, H. C. (1995) Characterization of the Gene Encoding the Polymorphic Immunodominant Molecule, a Neutralizing Antigen of *Theileria parva*. *J Immunol, 155*(3), pp. 1370-1381.

Ueti, M. W., Mealey, R. H., Kappmeyer, L. S., White, S. N., Kumpula-McWhirter, N., Pelzel, A. M., Grause, J. F., Bunn, T. O., Schwartz, A., Traub-Dargatz, J. L., Hendrickson, A., Espy, B., Guthrie, A. J., Fowler, W. K. and Knowles, D. P. (2012) Re-emergence of the apicomplexan *Theileria equi* in the United States: elimination of persistent infection and transmission risk. *PLoS One, 7*(9), pp. e44713.

Uilenberg, G. (1981) *Theilerial species of domestic livestock. In Advances in the control of theileriosis,* Martinus Nijhoff Publishers.

Uilenberg, G. (1984) Tick-borne livestock diseases and their vectors. 2. Epizootiology of tick-borne diseases. *World Animal Health Review, 8*, pp. 8-19.

Uilenberg, G., Schreuder, B. E., Mpangala, C. and Tondeur, W. (1977) Studies on *Theileriidae* (Sporozoa) in Tanzania IX. Unidentified bovine Theileriae. *Tropenmed Parasitol, 28*(4), pp. 494-8.

Uilenberg, G., Schreuder, B. E. C., Silayo, R. S. and Mpangala, C. (1976) Studies on Theileriidae (Sporozoa) in Tanzania .4. Field Trial on Immunization against East Coast Fever (*Theileria Parva* Infection of Cattle). *Tropenmed Parasitol, 27*(3), pp. 329-336.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M. and Rozen, S. G. (2012) Primer3--new capabilities and interfaces. *Nucleic Acids Res, 40*(15), pp. e115.

Urquhart, G. M., Armour, J., Duncan, J.L., Dunn, A.M., and Jennings, F.W. (1996) *Veterinary Parasitology,* Second ed.*,* Blackwell Science Ltd.

Vajana, E., Barbato, M., Colli, L., Milanesi, M., Rochat, E., Fabrizi, E., Mukasa, C., Del Corvo, M., Masembe, C., Muwanika, V. B., Kabi, F., Sonstegard, T. S., Huson, H. J., Negrini, R., NextGen, C., Joost, S. and Ajmone-Marsan, P. (2018) Combining Landscape Genomics and Ecological Modelling to Investigate Local Adaptation of Indigenous Ugandan Cattle to East Coast Fever. *Front Genet, 9*, pp. 385.

Vosloo, W., Bastos, A. D., Sangare, O., Hargreaves, S. K. and Thomson, G. R. (2002) Review of the status and control of foot and mouth disease in sub-Saharan Africa. *Rev Sci Tech, 21*(3), pp. 437-49.

Vudriko, P., Okwee-Acai, J., Tayebwa, D. S., Byaruhanga, J., Kakooza, S., Wampande, E., Omara, R., Muhindo, J. B., Tweyongyere, R., Owiny, D. O., Hatta, T., Tsuji, N., Umemiya-Shirafuji, R., Xuan, X., Kanameda, M., Fujisaki, K. and Suzuki, H. (2016) Emergence of multi-acaricide resistant *Rhipicephalus* ticks and its implication on chemical tick control in Uganda. *Parasit Vectors, 9*, pp. 4.

Walker, A. R., Bouattour, A., Camicas, J-L., Estrada-Pena, A., Horak, I.G., Latif, A.A., Pegram, R.G. & Preston, P.M. (2003) *Ticks of Domestic Animals in Africa,* Bioscience Reports.

Wamuyu, L., Obanda, V., Kariuki, D., Gakuya, F., Makanda, M., Otiende, M. and Ommeh, S. (2015) Molecular Detection and Characterization of *Theileria* Infecting Wildebeest (*Connochaetes taurinus*) in the Maasai Mara National Reserve, Kenya. *Pathogens, 4*(3), pp. 626-38.

Watt, D. M. and Walker, A. R. (2000) Pathological effects and reduced survival in *Rhipicephalus appendiculatus* ticks infected with *Theileria parva* protozoa. *Parasitology Research, 86*(3), pp. 207-14.

Wesonga, F. D., Gachohi, J. M., Kitala, P. M., Gathuma, J. M. and Njenga, M. J. (2015) *Theileria parva* infection seroprevalence and associated risk factors in cattle in Machakos County, Kenya. *Trop Anim Health Prod, 47*(1), pp. 93-101.

Wickham, H. (2017) *tidyverse: Easily Install and Load the 'Tidyverse'. R package version 1.2.1.*, Available: https://CRAN.R-project.org/package=tidyverse 2019].

Wint, W. and Kiara, H. (2017) *Where should the Muguga cocktail be used? The distributions of Theileria parva, its hosts and vectors in Africa*: Environmental Research Group Oxford (ERGO) c/o Department of Zoology, South Parks Road, Oxford, OX1 3PS, UK and International Livestock Research Institute (ILRI) Old Naivasha Road, near Uthiru, Nairobi, Kenya 00100.

Woolhouse, M. E., Thumbi, S. M., Jennings, A., Chase-Topping, M., Callaby, R., Kiara, H., Oosthuizen, M. C., Mbole-Kariuki, M. N., Conradie, I., Handel, I. G., Poole, E. J., Njiiri, E., Collins, N. E., Murray, G., Tapio, M., Auguet, O. T., Weir, W., Morrison, W. I., Kruuk, L. E., Bronsvoort, B. M., Hanotte, O., Coetzer, K. and Toye, P. G. (2015) Co-infections determine patterns of mortality in a population exposed to parasite infection. *Sci Adv, 1*(2), pp. e1400026.

Yeoman, G. H. (1966a) Field vector studies of epizootic East Coast Fever. I. A quantitative relationship between *R. appendiculatus* and the epizooticity of East Coast fever. *Bull Epizoot Dis Afr, 14*(1), pp. 5-27.

Yeoman, G. H. (1966b) Field vector studies of epizootic East Coast Fever. II. Seasonal studies of *R. appendiculatus* on bovine and non-bovine hosts in East Coast Fever enzootic, epizootic and free areas. *Bull Epizoot Dis Afr, 14*(2), pp. 113-40.

Yeoman, G. H. and Walker, J. B. (1967) *The Ixodid Ticks of Tanzania. A Study of the Zoogeography of the Ixodidae of an East African Country,* Commonwealth Institute of Entomology, London: The American Society of Tropical Medicine and Hygiene.

Yewdell, J. W. (2006) Confronting complexity: real-world immunodominance in antiviral CD8+ T cell responses. *Immunity, 25*(4), pp. 533-43.

Young A.S., L. B. L., Newson R.M. (1981) *The Occurrence of a Theileriaparva Carrier State in Cattle From an East Coast Fever Endemic Area of Kenya. In: Irvin A.D., Cunningham M.P., Young A.S. (eds) Advances in the Control of Theileriosis. Current Topics in Veterinary Medicine and Animal Science, vol 14,* The Hague, Boston, London: Martinus Nijhoff Publishers.

Young, A. S., Brown, C. G., Burridge, M. J., Cunningham, M. P., Kirimi, I. M. and Irvin, A. D. (1973) Observations on the cross-immunity between *Theileria lawrencei* (Serengeti) and *Theileria parva* (Muguga) in cattle. *Int J Parasitol, 3*(6), pp. 723-8.

Young, A. S., Brown, C. G., Burridge, M. J., Grootenhuis, J. G., Kanhai, G. K., Purnell, R. E. and Stagg, D. A. (1978) The incidence of theilerial parasites in East African buffalo (*Syncerus caffer)*. *Tropenmed Parasitol, 29*(3), pp. 281-8.

Young, A. S., Grootenhuis, J. G. and Irvin, A. D. (1985) *Influence of wildlife on immunization of cattle against theileriosis in East Africa. In: Irvin, A.D. (Ed), Immunization Against Theileriosis in Africa,* ILRAD, Nairobi.

Young, A. S., Leitch, B. L., Newson, R. M. and Cunningham, M. P. (1986) Maintenance of *Theileria parva parva* infection in an endemic area of Kenya. *Parasitology, 93 (Pt 1)*, pp. 9-16.

Young, A. S., Mutugi, J. J., Kariuki, D. P., Lampard, D., Maritim, A. C., Ngumi, P. N., Linyonyi, A., Leitch, B. L., Ndungu, S. G., Lesan, A. C. and et al. (1992) Immunisation of cattle against theileriosis in Nakuru District of Kenya by infection and treatment and the introduction of unconventional tick control. *Vet Parasitol, 42*(3-4), pp. 225-40.

Zweygarth, E., Just, M. C. and De Waal, D. T. (1997) In vitro cultivation of *Babesia equi:* detection of carrier animals and isolation of parasites. *Onderstepoort J Vet Res, 64*(1), pp. 51-6.

Zweygarth, E., Koekemoer, O., Josemans, A. I., Rambritch, N., Pienaar, R., Putterill, J., Latif, A. and Potgieter, F. T. (2009) *Theileria*-infected cell line from an African buffalo (*Syncerus caffer*). *Parasitology Research, 105*(2), pp. 579-81.

Appendix A: Certification of Questionnaire Ethical Approval

ROYAL (DICK) SCHOOL OF VETERINARY STUDIES
The University of Edinburgh
Easter Bush Veterinary Centre
Midlothian
EH25 9RG

R(D)SVS Human Ethical Review Committee

Reference: HERC_00_16
Project Title: **East Coast Fever and vaccination at the livestock/wildlife interface in Tanzania – questionnaire survey of livestock keepers.**
Lead Investigator: Fiona Allan

Status: Approved
Date: 05 January 2016

Signed:

Professor Susan Rhind
On behalf of HERC

Appendix B: Questionnaire

Questionnaire for Northern Tanzanian farmers in Serengeti National Park sampling sites

As part of the PhD of Fiona K Allan, The Roslin Institute, University of Edinburgh

Questionnaire for Farmers in SNP sampling sites

Date

Herd/farm identification number

GPS location of farm

Name of participant

……

Head of household? Yes No

If not head of household, please state relationship to head of household

……

……

1.a) How many cattle do you have in your herd **here now**?

b) Do all of the cattle in your herd belong to you? Yes No

c) If they do not all belong to you, how many belong to other people?

d) Are most of your cattle bred on this farm or are most of them bought in?

e) Do you keep your cattle at your farm here overnight? Yes ☐ No ☐

If no, where are they kept?.....

f) Where do you take your cattle to **graze** and **water**? (Please mark)

	Wet season	Dry season
Where do your cattle **graze**? Location: Time to get there: Distance to get there:		
Do you bring your cattle back to the farm every night? (Yes/No)		
Where do your cattle **water**? Location: Time to get there: Distance to get there:		
Do you bring your cattle back to the farm every night? (Yes/No)		

g) Do you ever send your cattle away for a period of time to a relative/neighbour?

Yes ☐ No ☐

- Do you ever send your cattle away for a period of time to a new area?

Yes ☐ No ☐

If yes, how long for?.....

Which months?.....

Location (place name):……

How far away (kms)?.....

Why?.....

h) **How many** other animals are on this farm? Sheep ☐ Goats ☐

2.a) Can you identify a **tick**? (please mark box)

b) Do you see ticks on your cattle? ☐ Yes ☐ No

c) On which part(s) of the body do you see ticks?……..

……

d) In which months do you see ticks? (please circle)

Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec

e) Which **areas** around your farm (or grazing & watering areas that you use) are ticks most commonly seen?.....

……

……..

Around house ☐ When grazing ☐ When watering ☐

When travelling to graze/water ☐ Everywhere ☐

f) Do you see ticks on the ears? ☐ Yes ☐ No

g) If yes, which **areas** around your farm (or grazing & watering areas that you use) do cattle pick up ticks **on the ears?**.....

……

……..

If yes, what time of year do you see most ticks on the **ears**?.....

……

……..

h) Do you know what diseases are spread by ticks?.....

……

……

3.a) Do you do anything to prevent ticks on your cattle? Yes ☐ No ☐

- b) If yes, what do you do? – (please mark as many as apply and mark when used)

Avoid areas where there are ticks when grazing	Hand removal	Use products to prevent ticks	Use products when you see ticks	Other (please state)

All year round					
Only when ticks are bad (state months)					

- c) If you use products, how do you apply the product(s)? – (please mark box)

Hand spray/pump (which part of body)	Brush	Dip tank	Pour-on (which part of body)	Other (please state)

<u>If spray/pump</u>

d) What is the time interval between treatments?.....

……

……

e) Product name (if known)……

Take photo of product if available.

f) What dilution do you use the product i.e. product:water ratio?.....

……

……

g) How much of the product do you use in total (for all the cows)? (please indicate if this dose is before or after dilution).....

……

……

h) Where do you buy the product(s)?.....

i) What do the product(s) cost?.....

<u>If dip</u>

j) What is the time interval between treatments?.....

……

……

k) Product name (if known)……

Take photo of product if available.

l) What dilution do you use the product i.e. product:water ratio?.....

……

……

m) How much of the product do you use in total (for all the cows)? (please indicate if this dose is before or after dilution).....

……

……

n) Where do you buy the product(s)?.....

o) What do the product(s) cost?.....

<u>If pour-on</u>

p) What is the time interval between treatments?.....

……

……

q) Product name (if known)……

Take photo of product if available.

r) What dilution do you use the product i.e. product:water ratio?.....

……

……

s) How much of the product do you use in total (for all the cows)? (please indicate if this dose is before or after dilution).....

……

……

t) Where do you buy the product(s)?.....

u) What do the product(s) cost?.....

v) Does the product just protect against ticks, or does it bring any other benefits?.....……..

……

……

4.a) Can you identify a **tsetse fly**? (please mark box)

b) Do you see tsetse flies near your cattle? Yes ☐ No ☐

c) In which months do you see tsetse flies?: (please circle)

Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec

d) Which **areas** around your farm (or grazing & watering areas that you use) are tsetse most commonly seen?……

……

……

Around house ☐ When grazing ☐ When watering ☐

When travelling to graze/water ☐ Everywhere ☐

e) Do you know what diseases are spread by tsetse?.....

……

……

f) Which cause more problems for your cattle – ticks or tsetse?.....

Explain……..

……

……

5.a) Do you do anything to prevent tsetse biting your cattle? Yes ☐ No ☐

- b) If yes, what do you do? – (please mark as many as apply and mark when used)

Avoid areas where there are tsetse when grazing	Use products to prevent tsetse	Products I use for ticks also protect against tsetse	Other (please state)

All year round				
Only when tsetse are bad (state months)				

- c) If you use products, how do you apply the product(s)? – (please mark box)

Hand spray/pump (which part of body)	Brush	Dip tank	Pour-on (which part of body)	Other (please state)

<u>If spray/pump</u>

d) What is the time interval between treatments?.....

……

……

e) Product name (if known)……

Take photo of product if available.

f) What dilution do you use the product i.e. product:water ratio?.....

……

g) What dose of product do you use in total (for all the cows)? (please indicate if this dose is before or after dilution).....

……

h) Where do you buy the product(s)?.....

i) What do the product(s) cost?.....

<u>If dip</u>

j) What is the time interval between treatments?.....

……

……

k) Product name (if known)……

Take photo of product if available.

l) What dilution do you use the product i.e. product:water ratio?.....

……

m) What dose of product do you use in total (for all the cows)? (please indicate if this dose is before or after dilution).....

……

n) Where do you buy the product(s)?.....

o) What do the product(s) cost?.....

<u>If pour-on</u>

p) What is the time interval between treatments?.....

……

……

q) Product name (if known)……

Take photo of product if available.

r) What dilution do you use the product i.e. product:water ratio?.....

……

s) What dose of product do you use in total (for all the cows)? (please indicate if this dose is before or after dilution).....

……

t) Where do you buy the product(s)?.....

u) What do the product(s) cost?.....

6.a) Have you heard of East Coast Fever? ☐ Yes ☐ No

b) If yes, do you know what the signs of East Coast Fever are? Yes ☐ No ☐

c) If yes, please mark the signs:

Elevated body temperature
Enlarged lymph nodes
Weight loss
Anaemia
Diarrhoea
Rough hair coat
Cough
Others (please state)

d) Do you know what causes East Coast Fever?..... Don't know ☐

e) Have any of your cattle had East Coast Fever?

☐ Yes ☐ No ☐ Don't know

f) In the last 1 year,

- How many **cases** of East Coast Fever? ☐

- How many **deaths** from East Coast Fever? ☐

g) Do you have methods to prevent East Coast Fever? Yes ☐ No ☐

h) If yes, please list these methods……...

……

……

i) Do you use the vaccination to protect against East Coast Fever?

Yes ☐ No ☐

j) If not, why not? (please mark box)

Did not know there was a vaccine	Too expensive	Does not work	Do not know where to buy it	Other (please state)

k) How do you treat cattle you suspect are sick with East Coast Fever? (state preferred treatment).....

……

……

l) Where do you get treatment from?.....

……

7.a) Have you heard of animal Trypanosomiasis? ☐ Yes ☐ No

b) If yes, do you know what the signs of animal Trypanosomiasis are?

Yes ☐ No ☐

c) If yes, please mark the signs:

Elevated body temperature
Enlarged lymph nodes
Weight loss
Anaemia
Diarrhoea
Rough hair coat
Cough
Others (please state)

d) Do you know what causes animal Trypanosomiasis?.....

Don't know ☐

e) Have any of your cattle had Trypanosomiasis? Yes ☐ No ☐

f) In the last 1 year,

- How many **cases** of Trypanosomiasis? ☐
- Home many **deaths** from Trypanosomiasis? ☐

g) Do you have methods to prevent animal Trypanosomiasis?

Yes ☐ No ☐

h) If yes, please list these methods ……..

……..……..

i) How do you treat cattle you suspect are sick with Trypanosomiasis? (State preferred order)…….

……

……

j) Where do you get treatment from?.....

……

Thank you for taking the time to complete this questionnaire, your time and input is greatly appreciated and the valuable information you have given will help to better understand the diseases mentioned.

Appendix C: Herd level risk factor analysis

Table C.0:1: Herd-level factors associated with *T. parva* prevalence based on logistic regression

Variable	Factor level	Total	*T. parva* positive (%)	OR	95% CI	p-value
Herd size	4-14	160	9/160 (5.63%)	REF		
	14-25	154	8/154 (5.19%)	0.92	0.34-2.47	0.866
	25-38	160	9/160 (5.63%)	1.00	0.38-2.63	1.000
	38-58	154	6/154 (3.89%)	0.68	0.22-1.93	0.475
	58-280	140	7/140 (5.00%)	0.88	0.31-2.43	0.810
Village location	Bwitengi	113	7/113 (6.19%)	REF		
	Bisarara	81	2/81 (2.47%)	0.38	0.06-1.64	0.240
	Makundusi	115	6/115 (5.22%)	0.83	0.26-2.59	0.751
	Nyiberekera	91	6/91 (6.59%)	1.07	0.33-3.33	0.908
	Park Nyigoti	92	3/92 (3.26%)	0.51	0.11-1.89	0.340
	Robanda	88	0/88 (0.00%)	$1.31x10^{-7}$	$6.34x10^{-100}$– $2.14x10^{7}$	0.982
	Singisi	114	10/114 (8.77%)	1.46	0.54-4.15	0.463
	Tamkeri	77	5/74 (6.49%)	1.05	0.30-3.42	0.934
Cattle all owned by participant	No	290	13/290 (4.48%)	REF		
	Yes	478	26/478 (5.44%)	1.23	0.63-2.49	0.559
Cattle sent away	No	648	29/648 (4.48%)	REF		
	Yes	120	10/120 (8.33%)	1.94	0.88-3.97	0.082
Sheep on farm	Sheep	691	35/691 (5.07%)	REF		
	No sheep	77	4/77 (5.19%)	1.03	0.30-2.66	0.961
Goats on farm	No goats	203	10/203 (4.93%)	REF		
	Goats	565	29/565 (5.13%)	1.04	0.52-2.29	0.908
Ticks seen on cattle	No	0	0/0 (0.00%)	REF		
	Yes	768	39/768 (5.08%)			
Ticks seen on ears	Ticks seen	656	32/624 (4.88%)	REF		
	Ticks not seen	112	7/112 6.25%)	1.29	0.52-2.86	0.542
Tick prevention	Yes	760	38/760 (5%)	REF		
	No	8	1/8 (12.5%)	2.71	0.14-15.81	0.356

Variable	Factor level	Total	*T. parva* positive (%)	OR	95% CI	p-value
Farmers know that ticks spread ECF	Farmers know	122	5/122 (4.09%)	REF		
	Farmers do not know	600	30/600 (5.00%)	1.23	0.51-3.67	0.673
ECF prevention methods	No prevention	42	1/42 (2.38%)	REF		
	ECF prevention	96	6/96 (6.25%)	4.07	0.67-77.82	0.199
ECF cases	Cases	42	1/42 (2.38%)	REF		
	No cases	150	7/15 (4.67%)	2.01	0.34-38.08	0.520
ECF deaths	No deaths	170	7/170 (4.11%)	REF		
	Deaths	22	1/22 (4.54%)	1.11	0.06-6.68	0.925
Spraying	Yes	551	28/551 (5.08%)	REF		
	No	217	11/217 (5.07%)	1.00	0.50-2.14	0.994
Spray time interval (days)$^{+}$	0-4	8	0/8 (0.00%)	REF		
	5-7	193	8/193 (4.15%)	$5.00x10^{6}$	$4.72x10^{-63}$- NA	0.994
	8-14	194	10/194 (5.15%)	5.99×10^{6}	$8.25x10^{-78}$–NA	0.995
	15-28	20	0/20 (0.00%)	1.00	$4.22x10^{-29}$– $2.37x10^{28}$	1.000
	29-50	90	6/90 (6.66%)	9.13×10^{6}	$9.05x10^{-49}$–NA	0.994
	>50	20	2/20 (10%)	1.28×10^{7}	$1.27x10^{-48}$-NA	0.994
Dipping	Yes	301	15/301 (4.98%)	REF		
	No	467	24/467 (5.14%)	1.03	0.54-2.04	0.924
Dip time interval (days)$^{+}$	7-14	185	13/185 (7.03%)	REF		
	14-30	44	0/44 (0.00%)	4.21 x 10-8	NA–$8.76x10^{40}$	0.991
	30-150	57	1/57 (1.75%)	0.23	0.01-1.18	0.159
Distance sent away (km)	0-7km	67	4/67 (5.97%)	REF		
	7-15km	60	6/60 (10%)	1.75	0.48-7.15	0.405
Distance grazing wet (km)	0.05-0.4	264	11/264 (4.16%)	REF		
	0.4-1	250	10/250 (4.00%)	0.96	0.39-2.31	0.924
	1-6	254	18/254 (7.09%)	1.75	0.82-3.91	0.153
Distance grazing dry (km)	0.05-2	300	16/284 (5.33%)	REF		

Variable	Factor level	Total	*T. parva* positive (%)	OR	95% CI	p-value
	2-5	295	16/295 (5.42%)	1.02	0.49-2.09	0.961
	5-10	173	7/173 (4.05%)	0.75	0.29-1.79	0.532
Distance watered wet (km)	0.1-0.5	258	13/258 (5.04%)	REF		
	0.5-2	428	21/428 (4.91%)	0.97	0.48-2.03	0.938
	2-3	82	5/82 (6.09%)	1.22	0.38-3.36	0.710
Distance watered dry (km)	0.3-1.5	278	15/278 (5.39%)	REF		
	1.5-3.5	237	11/237 (4.64%)	0.85	0.37-1.88	0.697
	3.5-6	253	13/253 (5.14%)	0.95	0.44-2.04	0.895
Farmers know of ECF	Yes	312	13/312 (4.16%)	REF		
	No	456	26/456 (5.70%)	1.39	0.72-2.83	0.343
Distance from boundary[1] (km)	0-1.42	138	4/138 (2.89%)	REF		
	1.42-2.56	135	8/135 (5.93%)	2.11	0.65-8.06	0.232
	2.56-3.48	131	3/131 (2.29%)	0.79	0.15-3.63	0.755
	3.48-5.37	116	7/116 (6.03%)	2.15	0.63-8.39	0.231
	5.37-7.87	123	9/123 (7.32%)	2.64	0.84-9.97	0.113
	7.87-9.49	128	8/128 (6.25%)	2.23	0.69-8.54	0.199
Ticks seen in certain season	Oct – May (wet)	132	4/132 (3.03%)	REF		
	June – Sept (dry)	639	35/639 (5.48%)	1.85	0.72-6.29	0.25
Product used	Albadip	354	17/354 (4.80%)	0.89	0.17-3.59	0.878
	Paranex	206	10/206 (4.85%)	0.90	0.17-3.70	0.893
	Cybadip	92	5/92 (5.43%)	0.88	0.18-2.91	0.851
	Tantix	75	2/75 (2.67%)	0.42	0.05-2.10	0.354

1 The distance from each farm to the boundary of the protected area

\+ Due to complete separation, caution must be taken in interpreting variables with zero prevalence

Appendix D: Tick sampling data

In total, 38 adult ticks were captured, over a two week period sampling several times daily. *Amblyomma* were numerically dominant, with 26 adults of 38 total ticks. The number of *Rhipicephalus appendiculatus* was very low, with only 2 adults of 38 total ticks captured. There was no noticeable difference in tick capture in the sweep net compared to the drag blanket. Table D.0:1 shows tick capture results from Ikorongo Game Reserve, Serengeti National Park, various farming sites around the SENAPA area, Singita concession and Ikoma open area.

Table D.0:1: Tick capture data from protected areas and farming areas

Site	GPS	Elevation (m)	Vegetation	Adult Ticks caught
Ikorongo Game Reserve				
1	S 01.99389 E 034.66276	1423	Long grass, very wet	None
2	S 01.99679 E 034.66704	1413	Long grass, very wet	None
3	S 01.99801 E 034.66840	1404	Very long, wet grass. 1 buffalo few hundred metres away	None
4	S 02.00191 E 034.67377	1397	Very long grass, damp. Some buffalo dung	1 – unknown
5	S 02.01078 E 034.68461	1396	Very long grass, drier	None
6	S 02.02543 E 034.69228	1387	Grass long, drier	1 – unknown
7	S 02.02720 E 034.68952	1388	Shrubby shorter grass, drier	1 - *R. appendiculatus*
Serengeti National Park				
8	S 02.41546 E 034.89140	1548	Dry long grass (above knee), shrubby trees. Hot and dry. 3 buffalo, 1 topi in vicinity	5 – *Amblyomma*
9	S 02.42275 E 034.89154	1542	Shorter grass, wetter. Mud hole nearby. Hot and dry. 2 buffalo, 1 ostrich, 1 grants gazelle, 1 impala	2 - *Amblyomma*
10	S 02.42764 E 034.88752	1413	Long grass, very wet with dew. Mud hole nearby. 10 buffalo nearby	None
11	S 02.43029 E 034.88671	1556	Knee-length grass, not so wet. Batchelor herd of ~30 buffalo nearby	1 - *Amblyomma*
12	S 02.43149 E 034.88480	1552	Longer grass. Lots of buffalo nearby	1 – unknown
13	S 02.43233 E 034.88391	1552	Grass waist-height but sparse, wet. Buffalo just risen from sleep	2 - *Amblyomma*
Farming areas 12/02/2016				
14	S 02.05215	1265	Short grass, damp	None

Site	GPS	Elevation (m)	Vegetation	Adult Ticks caught
	E 034.37566			
15	S 01.97631 E 034.33483	1259	Short shrubby grass, drier	None
Farming areas 17/02/2016				
16	S 01.96558 E 034.55822	1375	Grass calf-length, dry. Cattle around	None
17	S 01.96259 E 034.56129	1385	Grass calf-length, dry	None
18	S 01.96051 E 034.62999	1458	Grass short, shrubby.	None
Farming areas 19/02/2016				
19	S 01.82505 E 034.81598	1620	Short grass, shrubby and over-grazed. Dry. Cattle nearby	None
20	S 01.82500 E 034.81609	1635	Short grass, dry	None
21	S 01.79523 E 034.76694	1629	Grass calf-length, dry	None
Singita Concession area				
22	S 2.085827 E 34.387936	1414	Grass knee-length, trees interspersed. Dry	13 – mixed
Ikoma open area				
23	S 2.069091 E 34.615528	1429	Grass knee-length, bushes interspersed. Wildlife known to pass through	1 – *R. appendiculatus* 10 – *Amblyomma*

Appendix E: PacBio long read sequencing results

Table E.0:1: Tp1 read counts for each individual sample at each stage of filtering

	Individual ID	CCS Reads	Full length/size/No. passes filter
	TpM	8,591	3,400
Buffalo	buffalo|Serengeti|2011|1749	3,793	45
	buffalo|Serengeti|2011|1751	10,538	2,630
	buffalo|Serengeti|2011|1757	9,944	308
	buffalo|Serengeti|2011|1763	13,533	720
	buffalo|Serengeti|2011|1765	11,336	3,978
	buffalo|Serengeti|2011|1769	9,451	1,258
	buffalo|Serengeti|2011|1775	10,189	1,206
	buffalo|Serengeti|2011|1777	9,749	1,070
	buffalo|Serengeti|2011|1781	6,264	1,890
	buffalo|Serengeti|2011|1783	8,030	3,280
	buffalo|Serengeti|2011|1787	6,728	2,754
	buffalo|Serengeti|2011|1789	13,679	29
Cross-sectional 2016 cattle	cattle|cross-sect_BW01|2016|31	9,943	3,902
	cattle|cross-sect_BW02|2016|19	6,941	2,756
	cattle|cross-sect_BW05|2016|103	10,368	4,081
	cattle|cross-sect_BW09|2016|53	7,656	3,246
	cattle|cross-sect_BW09|2016|71	12,080	4,657
	cattle|cross-sect_MA03|2016|472	18,337	5,225
	cattle|cross-sect_MA07|2016|529	14,137	5,902
	cattle|cross-sect_NY07|2016|750	1,756	697
	cattle|cross-sect_PA07|2016|350	5,528	2,143
	cattle|cross-sect_SI02|2016|574	8,452	3,540
	cattle|cross-sect_SI08|2016|627	1,413	573
	cattle|cross-sect_SI08|2016|628	6,580	2,742
	cattle|cross-sect_TA02|2016|195	9,031	3,534
	cattle|cross-sect_TA02|2016|201	4,125	1,566
Cross-sectional 2011 cattle	cattle|cross-sect_Iserere|2011|6429	6,201	2,424
	cattle|cross-sect_Iserere|2011|6463	0	0
	cattle|cross-sect_Iserere|2011|6464	9,650	3,889
	cattle|cross-sect_Kenyamonta|2011|6403	14,347	5,987
	cattle|cross-sect_Masangura|2011|6226	8,007	3,176
	cattle|cross-sect_Nyamsingisi|2011|6107	11,939	4,362
	cattle|cross-sect_Nyamsingisi|2011|6120	5,953	2,346
	cattle|cross-sect_Ringwani|2011|6305	3,102	1,282
	cattle|cross-sect_Ringwani|2011|6310	4,107	1,493

	Individual ID	CCS Reads	Full length/size/No. passes filter
	cattle|cross-sect_Ringwani|2011|6317	10,546	2,874
	cattle|cross-sect_Ringwani|2011|6347	9,681	3,806
	cattle|cross-sect_Singisi|2011|6032	6,496	2,671
Longitudinal cattle	cattle|longitudinal_Nyamburi|2013|8180	13,190	5,363
	cattle|longitudinal_Nyamburi|2017|10507	11,213	3,934
	cattle|longitudinal_Nyamburi|2017|9202	8,437	3,491
	cattle|longitudinal_Nyichoka|2013|8459	3,362	1,258
	cattle|longitudinal_Nyichoka|2017|8450	13,449	5,193
	cattle|longitudinal_Rwamchanga|2017|9750	5,731	2,275
Clinically ill cattle	cattle|sick_Nyichoka|2014|7612	14,752	6,075
	cattle|sick_Nyichoka|2014|7628	6,423	2,498
	Total	**404,758**	**131,529**

Table E.0:2: N60 read counts for each individual sample at each stage of filtering

	Individual ID	CCS Reads	Full length/size/No. passes filter
	TpM	6,304	3,168
Buffalo	buffalo|Serengeti|2011|1747	3,453	880
	buffalo|Serengeti|2011|1749	7,513	813
	buffalo|Serengeti|2011|1751	9,727	954
	buffalo|Serengeti|2011|1753	9,526	1,070
	buffalo|Serengeti|2011|1755	7,433	833
	buffalo|Serengeti|2011|1757	6,217	699
	buffalo|Serengeti|2011|1759	9,696	844
	buffalo|Serengeti|2011|1761	7,568	594
	buffalo|Serengeti|2011|1763	10,890	988
	buffalo|Serengeti|2011|1765	9,178	598
	buffalo|Serengeti|2011|1767	6,325	323
	buffalo|Serengeti|2011|1769	6,388	550
	buffalo|Serengeti|2011|1771	9,184	1,454
	buffalo|Serengeti|2011|1773	8,515	901
	buffalo|Serengeti|2011|1775	11,265	832
	buffalo|Serengeti|2011|1777	9,128	631
	buffalo|Serengeti|2011|1781	8,603	814
	buffalo|Serengeti|2011|1783	6,317	529
	buffalo|Serengeti|2011|1785	10,028	943
	buffalo|Serengeti|2011|1787	8,371	4,206
	buffalo|Serengeti|2011|1789	10,870	1,076
Cross-sectional 2016 cattle	cattle|cross-sect_BI07|2016|168	0	0
	cattle|cross-sect_BW01|2016|31	10,748	5,490
	cattle|cross-sect_BW02|2016|19	6,055	3,021
	cattle|cross-sect_BW05|2016|103	7,437	3,636
	cattle|cross-sect_BW08|2016|97	726	384
	cattle|cross-sect_BW09|2016|53	8,252	4,239
	cattle|cross-sect_BW09|2016|71	7,939	4,026
	cattle|cross-sect_MA02|2016|481	3,806	1,801
	cattle|cross-sect_MA03|2016|472	5,691	2,784
	cattle|cross-sect_MA07|2016|517	4,821	2,320
	cattle|cross-sect_MA07|2016|529	7,352	3,770
	cattle|cross-sect_MA09|2016|547	341	137
	cattle|cross-sect_NY03|2016|717	3,698	1,682
	cattle|cross-sect_NY07|2016|747	0	0
	cattle|cross-sect_NY07|2016|750	3,389	1,683

	Individual ID	CCS Reads	Full length/size/No. passes filter
	cattle|cross-sect_NY09|2016|759	419	190
	cattle|cross-sect_PA07|2016|350	2,766	1,410
	cattle|cross-sect_PA07|2016|355	0	0
	cattle|cross-sect_SI01|2016|621	5,883	2,793
	cattle|cross-sect_SI02|2016|574	5,334	2,559
	cattle|cross-sect_SI08|2016|627	0	0
	cattle|cross-sect_SI08|2016|628	3,192	1,598
	cattle|cross-sect_SI09|2016|646	3,004	1,437
	cattle|cross-sect_TA02|2016|195	1,934	795
	cattle|cross-sect_TA02|2016|201	177	73
Cross-sectional 2011 cattle	cattle|cross-sect_Kenyamonta|2011|6365	0	0
	cattle|cross-sect_Kenyamonta|2011|6367	0	0
	cattle|cross-sect_Kenyamonta|2011|6371	0	0
	cattle|cross-sect_Kenyamonta|2011|6373	0	0
	cattle|cross-sect_Kenyamonta|2011|6397	0	0
	cattle|cross-sect_Kenyamonta|2011|6401	0	0
	cattle|cross-sect_Kenyamonta|2011|6403	2,832	1,404
	cattle|cross-sect_Nyamsingisi|2011|6102	396	205
	cattle|cross-sect_Nyamsingisi|2011|6107	3,366	1,488
	cattle|cross-sect_Nyamsingisi|2011|6109	574	288
	cattle|cross-sect_Nyamsingisi|2011|6110	2,338	1,175
	cattle|cross-sect_Nyamsingisi|2011|6120	5,844	2,963
	cattle|cross-sect_Ringwani|2011|6305	7,176	3,591
	cattle|cross-sect_Ringwani|2011|6310	1,143	581
	cattle|cross-sect_Ringwani|2011|6315	11,540	5,853
	cattle|cross-sect_Ringwani|2011|6317	10,509	5,361
	cattle|cross-sect_Ringwani|2011|6334	0	0
	cattle|cross-sect_Ringwani|2011|6343	113	0
	cattle|cross-sect_Ringwani|2011|6347	9,817	4,925
	cattle|cross-sect_Singisi|2011|5994	7,077	2,910
	cattle|cross-sect_Singisi|2011|5996	8,274	4,162
	cattle|cross-sect_Singisi|2011|5997	1,057	514
	cattle|cross-sect_Singisi|2011|5998	0	0
	cattle|cross-sect_Singisi|2011|5999	6,103	3,073
	cattle|cross-sect_Singisi|2011|6014	139	60
	cattle|cross-sect_Singisi|2011|6016	132	64
	cattle|cross-sect_Singisi|2011|6021	1,355	652
	cattle|cross-sect_Singisi|2011|6027	0	0
	cattle|cross-sect_Singisi|2011|6032	4,172	2,063
	cattle|cross-sect_Singisi|2011|6036	619	283

	Individual ID	CCS Reads	Full length/size/No. passes filter
	cattle|cross-sect_Singisi|2011|6038	2,151	1,058
	cattle|cross-sect_Singisi|2011|6226	8,109	3,989
Longitudinal cattle	cattle|longitudinal_Nyamburi|2013|8180	10,244	5,163
	cattle|longitudinal_Nyamburi|2015|9202	7,816	3,885
	cattle|longitudinal_Nyamburi|2015|9215	0	0
	cattle|longitudinal_Nyamburi|2017|10507	7,894	4,011
	cattle|longitudinal_Nyamburi|2017|9202	0	0
	cattle|longitudinal_Nyichoka|2013|8446	0	0
	cattle|longitudinal_Nyichoka|2013|8458	6,522	3,237
	cattle|longitudinal_Nyichoka|2013|8459	3,467	1,754
	cattle|longitudinal_Nyichoka|2013|8460	0	0
	cattle|longitudinal_Nyichoka|2015|7597	161	58
	cattle|longitudinal_Nyichoka|2015|9341	0	0
	cattle|longitudinal_Nyichoka|2017|8450	2,751	0
	cattle|longitudinal_Nyichoka|2017|9671	0	0
	cattle|longitudinal_ParkNyigoti|2015|8479	4,878	2,256
	cattle|longitudinal_Rwamchanga|2017|9750	7,371	3,632
Clinically ill cattle	cattle|sick_Nyichoka|2014|7612	7,893	3,604
	Total	**429,296**	**143,790**